

ACCURATE PORTFOLIO RISK-RETURN STRUCTURE MODELLING



A Dissertation presented
by

Nafees Hossain

to

The Department of Statistical Science

in fulfillment of the requirements
for the degree of

Doctor of Philosophy

in the subject of

Statistics

University of Cape Town
October 2006

The copyright of this thesis vests in the author. No quotation from it or information derived from it is to be published without full acknowledgement of the source. The thesis is to be used for private study or non-commercial research purposes only.

Published by the University of Cape Town (UCT) in terms of the non-exclusive license granted to UCT by the author.

Declaration

I Nafees Muzaffar Hossain hereby

1. grant the University of Cape Town free licence to reproduce the above thesis in whole or in part, for the purpose of research;
2. declare that:
 - the above thesis is my own unaided work, both in concept and execution, apart from the normal guidance from my supervisors.
 - neither the substance nor any part of the above thesis has been submitted in the past, or is being, or is to be submitted for a degree at this University or at any other university.

I am now presenting the thesis for examination for the degree of PhD.

SIGNED:

DATE:

Acknowledgments

I wish to express my sincere thanks to the following:

- In particular to my supervisors Professor Casper Troskie and Professor Renkuan Guo for always making time to answer my many questions, supplying me with many good ideas, for all the data used in this thesis, supplying useful material regarding the current study, and encouraging me to pursue further studies.
- To Professor Dunne for all his support and encouragement.
- To Allan Clark who often helped me with statistical and computing issues when I was experiencing difficulties.
- To all my class mates, in particular my good friend John Kyeyune, who assisted me during the course of the degree with numerous obstacles.
- To my father, Dr. M.M. Hossain for all his encouragement and important advice throughout my academic career and for the colossal love and support he has shown in my life.
- To my mother, Mrs. N.S. Hossain for all her encouragement, love and support and for always worrying about my health and studies.
- To my friends for their continued support and love.
- To the National Research Foundation and UCT for providing support during my studies.

Abstract

Markowitz's modern portfolio theory has played a vital role in investment portfolio management, which is constantly pushing the development on volatility models. Particularly, the stochastic volatility model which reveals the dynamics of conditional volatility. Financial time series and volatility models has become one of the hot spots in operations research. In this thesis, one of the areas we explore is the theoretical formulation of the optimal portfolio selection problem under Itô calculus framework. Particularly, a stochastic variation calculus problem, i.e., seeking the optimal stochastic volatility diffusion family for facilitating the best portfolio selection identified under the continuous-time stochastic optimal control theoretical settings. One of the properties this study examines is the left-shifting role of the $GARCH(1, 1)$ (General Autoregressive Conditional Heteroskedastic) model's efficient frontier. This study considers many instances where the left shifting superior behaviour of the $GARCH(1, 1)$ is observed. One such instance is when $GARCH(1, 1)$ is compared within the volatility modelling extensions of the $GARCH$ environ in a single index framework. This study will demonstrate the persistence of the superiority of the $GARCH(1, 1)$ frontier within a multiple and single index context of modern portfolio theory.

Many portfolio optimization models are investigated, particularly the Markowitz model and the Sharpe Multiple and Single index models. The Sharpe Models are based on the first two moments of the random components with uncorrelated variance-covariance structure.

This thesis gives rise to a new formulation with regards to index models by introducing a correlated variance-covariance structure, i.e. the Troskie-Hossain model otherwise referred as the Improved Sharpe framework. The focus of this study is to examine and compare the behaviour of efficient frontiers in a single and multiple index setting for the Sharpe Index and the Improved Sharpe Index portfolio models. The risk-return structures are explored under 4 estimation methods namely, $GARCH(1, 1)$, $ARMA$ (Autoregressive and Moving Average), Regression and State Space models i.e. the Kalman Filter. The exploration of the left-shifting role of the $GARCH(1, 1)$ model is investigated in the Sharpe and Improved Sharpe frameworks. The domineering behaviour of $GARCH(1, 1)$ is explored by means of comparisons set against the aforementioned estimation methods.

An application of Principal Components Analysis is considered for constructing significant orthogonal components of indices in order to attain efficient frontiers that illustrate an accurate risk-return structure for a portfolio of stocks on the South African stock exchange.

Finally we attempt to design dynamic simulated models for creating realistic sample paths of stock index returns from which 9 positively correlated stock returns are to be simulated via Monte-Carlo methods. The stock market indices are simulated using Geometric Brownian motion models from which we construct a simulated portfolio of asset values. Adopting appropriate Monte-Carlo methods the portfolio of stocks will be simulated exhibiting a $GARCH(1, 1)$ volatility process. A comparison is then conducted on the behaviour of the simulated portfolio under the Sharpe Multiple Index and the Improved Sharpe Multiple Index portfolio models.

Contents

Abstract	1
Contents	3
1 Introduction	13
1.1 Background to the Problem	13
1.2 Thesis Contributions	18
1.3 Objectives of the Study	20
1.4 Limitations of the Study	21
1.5 Plan of Development of the thesis	22
2 Some Fundamentals of Stock Market Theory	28
2.1 Introduction	28
2.2 Stationarity and Non-Stationarity	29
2.3 Some Financial Stock Theory: Fundamental Concepts	32
2.3.1 Distributional Properties of Stock Returns	37
2.4 The Univariate Normal Distribution. Tests for Normality.	39
2.4.1 The Multivariate Normal Distribution.	41
2.4.2 The Conditional Multivariate Normal Distribution	44
2.5 The Correlation and Autocorrelation Functions.	47
2.5.1 The Partial Autocorrelation Function	49
2.5.2 Portmanteau Test	51

2.6	White Noise and Linear Time Series.	52
2.7	Validity of the Normal Assumption for forthcoming Analysis.....	54
2.8	List Of Appendices	56
2.8.1	Appendix 1: Histogram of Monthly Log Returns of the remaining Eight Shares.....	56
3	Applied Regression Analysis and Model Selection	57
3.1	Introduction.....	57
3.2	Applied Regression Analysis	59
3.2.1	Assumptions underlying the Multiple Linear Regression Model. ...	60
3.2.2	Method of Moments and Least Squares	61
3.2.3	Properties of the OLS Estimators in the Multiple Regression Model	64
3.2.4	Properties of the OLS Estimators in the Simple Regression Model..	65
3.2.5	All Subsets Regression.....	66
3.3	Description of Model selection Methods and Criteria.....	68
3.3.1	Stepwise Regression	68
3.3.2	Forward Selection Procedure.	69
3.3.3	Backward Selection Procedure.....	73
3.3.4	The Cp criterion of Mallows.....	74
3.3.5	Information Criteria and Complexity.	75
3.3.6	The Model Building Process	76
3.4	Empirical Study: Single Index	82
3.4.1	Introduction	82
3.4.2	The data.....	83
3.4.3	Study Objectives	84

3.4.4	Methodology	84
3.4.5	Primary Findings.....	85
3.4.6	Conclusions	87
3.5	Empirical Study: Multiple Index	88
3.5.1	The data.....	88
3.5.2	Study Objectives	88
3.5.3	Methodology	89
3.5.4	Primary Findings.....	90
3.5.5	Conclusions	93
3.6	List Of Appendices	93
3.6.1	Appendix 1A: The Complete Models: Summary of Beta coefficients	94
3.6.2	Appendix 1B: The Final Models: Summary of Beta coefficients	95
4	Linear Time Series Models and Empirical Studies	96
4.1	Introduction.....	96
4.2	Objectives of Time Series analysis	98
4.3	Autoregressive Models	100
4.3.1	Identifying $AR(p)$ Models in Practice.....	113
4.3.2	The $AR(p)$ Process and Regression Analysis.	114
4.3.3	Estimation.....	117
4.3.4	Model Checking.....	118
4.4	Empirical Study: Single Index	119
4.4.1	Introduction.....	119
4.4.2	The Data	120

4.4.3	Study Objectives	120
4.4.4	Methodology	121
4.4.5	Primary Findings	122
4.4.6	Conclusions	126
4.5	List of Appendices	128
4.5.1	Appendix 2A: Correlograms for the residuals from the regression models for the remaining eight stocks	128
4.5.2	Appendix 2B: Correlograms for the residuals from the Time Series models for R_3, R_4, R_8 and R_9 stocks	129
4.6	Empirical Study: Multiple Index	130
4.6.1	The Data	130
4.6.2	Study Objectives	130
4.6.3	Methodology	130
4.6.4	Primary Findings	131
4.6.5	Conclusions	135
4.7	List of Appendices	136
4.7.1	Appendix 2C: Correlograms for the residuals from the 9 MIR time series models	136
4.7.2	Appendix 2D: The MIR time series Models: Summary of Beta coefficients	137
4.7.3	Appendix 2E: Correlograms of the Final MIR- time series Models for R_3, R_4, R_8 and R_9	138
5	State Space Modelling and Empirical Studies	139
5.1	Introduction	139
5.2	Representation and Estimation	140
5.3	The Kalman Filter	142

5.3.1	Introduction	142
5.3.2	The Kalman Filter Model	143
5.3.3	Recursive Estimation of the Model	144
5.3.4	The Kalman Filtering Process	150
5.3.5	Residuals and Maximum Likelihood Estimates	151
5.4	Empirical Investigation: Single Index	152
5.4.1	Introduction	152
5.4.2	The data	153
5.4.3	Study Objectives	153
5.4.4	Methodology	153
5.4.5	Primary Findings	154
5.4.6	Conclusions	157
5.5	Empirical Investigation: Multiple Index	158
5.5.1	The data	158
5.5.2	Study Objectives	158
5.5.3	Methodology	158
5.5.4	Primary Findings	159
5.5.5	Conclusions	160
5.6	List of Appendices	162
5.6.1	Appendix 3A: MI-state space coefficients	162
6	Conditional Heteroskedastic Models-Volatility Models	163
6.1	Introduction	163
6.2	The Characteristics of Volatility	165

6.3	Sources of ARCH and variables used to model Volatility	166
6.4	Structure of A Model	170
6.5	Different Specifications of the error distribution	173
6.5.1	ARCH Models with a t distribution	173
6.5.2	ARCH Models with a Skewed Standardised t distribution	174
6.6	Linear Regression with ARCH error terms	176
6.7	Other ARCH specifications	177
6.7.1	The ARCH in Mean Model	177
6.8	The weaknesses of the ARCH Model	178
6.9	The GARCH Model	179
6.10	The GARCH(1,1) Model	181
6.11	Limitations of the GARCH Model	182
6.12	Empirical Investigation: Single Index	183
6.12.1	The data	183
6.12.2	Study Objectives	183
6.12.3	Methodology	183
6.12.4	Primary Findings	184
6.12.5	Conclusions	187
6.13	Empirical Investigation: Multiple Index	188
6.13.1	The data	188
6.13.2	Study Objectives	188
6.13.3	Methodology	189
6.13.4	Primary Findings	190
6.13.5	Conclusions	192

6.14 List of Appendices	194
6.14.1 Appendix 4A: MI-Garch(1,1) Coefficients	194
6.15 A Summary of the ARCH and GARCH Models	195
6.16 Additional ARCH Models	197
6.16.1 The EGARCH Model	197
6.16.2 The Threshold GARCH (TARCH) Model	199
6.16.3 The Power ARCH (PARCH) Model	199
6.16.4 The Component GARCH (C-GARCH) Model	200
6.17 Empirical Investigation: Single Index	201
6.17.1 The data	201
6.17.2 Study Objectives	201
6.17.3 Methodology	202
6.17.4 Primary Findings	203
6.17.5 Conclusions	204
7 Modern Portfolio Theory	206
7.1 Classical Formulations	206
7.1.1 Efficient Frontier.	208
7.2 Modern Portfolio Theory: The Corner Portfolios of Sharpe.....	211
7.3 The Capital Market Line	218
7.3.1 Estimation of μ and Σ	219
7.4 The Index Model of Sharpe	220
7.5 The Single Index Model of Sharpe.....	223
7.6 Estimation of the parameters using Least Squares.	227

7.7 Empirical Study: Single Index	231
7.7.1 The Data	231
7.7.2 Study Objectives	231
7.7.3 Methodology	231
7.7.4 Primary Findings.....	232
7.7.5 Conclusion	233
7.8 GARCH(1,1) versus GARCH Extensions	233
7.8.1 The Data	234
7.8.2 Study Objectives	234
7.8.3 Methodology	234
7.8.4 Primary Findings.....	235
7.8.5 Conclusion	236
7.9 Generalisation of the Markowitz Formulations	236
7.10 The Multiple Index Model of Sharpe.	238
7.11 Empirical Study: Multiple Index	243
7.11.1 The Data	243
7.11.2 Study Objectives	243
7.11.3 Methodology	243
7.11.4 Primary Findings.....	244
7.11.5 Conclusion	245
8 The Troskie Hossain Innovations: The Improved Sharpe	246
8.1 Innovations to the Sharpe Index Model.....	246
8.2 The Improved Sharpe Single Index Model by Troskie and Hossain.	247

8.3 Improved Sharpe versus Sharpe - Single Index Case	251
8.3.1 The Data	251
8.3.2 Study Objectives	251
8.3.3 Methodology	251
8.3.4 Primary Findings	252
8.3.5 Conclusion	255
8.4 Improved Sharpe Multiple Index Model.	256
8.5 Improved Sharpe versus Sharpe - Multiple Index Case	260
8.5.1 The Data	260
8.5.2 Study Objectives	260
8.5.3 Methodology	261
8.5.4 Primary Findings	262
8.5.5 Conclusions	267
9 Principal Components Analysis	269
9.1 Introduction	269
9.2 Some Mathematical Background	272
9.3 The Singular Value Decomposition - SVD - of a Matrix.	278
9.4 Summary of Objectives and Benefits of PCA	285
9.5 Algorithm of PCA	286
9.6 Principal Components and Regression.	287
9.7 PCA Study: SMI vs ISMI - Multiple Index Case	288
9.7.1 The Data	288
9.7.2 Study Objectives	289

9.7.3 Methodology	290
9.7.4 Primary Findings.....	292
9.7.5 Conclusion	293
10 Monte Carlo Simulation Study	294
10.1 Introduction.....	294
10.2 Brownian Motion	296
10.2.1 Introduction.....	296
10.2.2 Einstein's Brownian Motion	297
10.2.3 Random Walk	298
10.2.4 Geometric Brownian Motion	301
10.2.5 Ito's Lemma.....	301
10.2.6 Application to Geometric Brownian Motion	302
10.3 Monte Carlo Simulation of Geometric Brownian Motion	303
10.4 The Methodology	305
10.5 Empirical Study.....	308
10.5.1 Conclusion	309
11 Summary And Directions For Further Research	310
References	313

Chapter 1

Introduction

1.1 Background to the Problem

Risk as measured by variance plays a major role in many financial applications. The (Capital Asset Pricing Model) CAPM or Market Model (Sharpe (1964), Lintner (1965), Mossin(1966) and Merton(1973)) for example posits a positive relationship between the expected returns of a share and the market portfolio. This hypothesis relies on the assumption of a constant market variance. Portfolio managers as well as traders thus require an estimate of the level of future market uncertainty (variance) if they believe in the validity of the Capital Asset Pricing Model (CAPM). Volatility Models, first developed by Engle (1982), model the conditional volatility (conditional variance) of time series.

Volatility models have become increasingly important since asset return variances as well as the covariances of asset returns were found not to be constant over time (Bollerslev (1986)) but rather that they evolve over time. Many researchers have used volatility models in order to model share returns, exchange rate movements as well as interest rate movements. Both univariate time series models as well as their multivariate extensions are important to portfolio managers. For example, since covariances are time varying one could use a multivariate volatility model in order to solve the asset allocation problem of Markowitz (1952). Volatility models have also been used in order to test certain economic theories. Bollerslev, Engle and Wooldridge (1988) used a trivariate CAPM in order to test

the CAPM and the assumption of constant covariances and variances amongst the three assets used (treasury bills, bonds and stocks). They found that the conditional covariances between the three assets were varied over time. It was also shown that this movement could be forecasted.

Currently volatility models are most often used by derivatives analysts since they rely on estimates of the variance (vols) of certain financial instruments in order to use these estimates as inputs into their asset pricing models. Banks and large corporates however also use volatility models in order to estimate their daily value at risk (VAR). From a statistical point of view volatility models are important since many commonly used tools are no longer valid in the presence of non-constant variances. For example, standard regression type models assume that the residuals from the regression model are homoskedastic (constant over time). If this assumption is violated the analyst should adjust any results from such a model in order to compensate for heteroskedastic errors. (Autoregressive and Conditional Heteroskedastic) ARCH Models in itself does not invalidate standard OLS inference. However, ignoring ARCH effects may result in the loss of efficiency of the estimated beta coefficients. From the preceding discussion it can be seen that volatility models are important to both the financial practitioner as well as to the practicing statistician. In what follows is a discussion on both the theory and the use of volatility models in the financial markets. Most of the references pertain to the South African stock market.

Volatility models were established in order to model the time varying nature of the conditional variance of time series. A number of researchers have tried to explain why conditional variances should be serially correlated. Diebold and Nerlove (1989) believe that

share returns are heteroskedastic due to the existence of a serially correlated news process. Gallant, Hsieh and Tauchen (1989) offer evidence in favour of the above hypothesis however Engle, Ito and Lin (1990a,b) were unable to provide any satisfactory explanation for the dependence in the underlying news arrival process.

Lamoureux and Lastrapes (1990b) argue that heteroskedasticity found in share returns are due to the clustering in trading volumes. When they included trading volumes into their variance equation they found that the lagged residuals were not significant thus substantiating their claim. Karpoff (1987) however showed that trading volumes and the price of a share are highly correlated. This could be the reason why Lamoureux and Lastrapes (1990b) found a significant loading on the trading volume in their variance equation.

In the quest to model the conditional variance of a share's return series, researchers have unearthed numerous economic variables that are related to the conditional variance of share returns. Gallant, Rossi and Tauchen (1992) found that lagged volume were positively related to the conditional volatility on the NYSE where as Campbell (1987) and Glosten, Jagannathan and Runkle (1991) found that nominal interest rates and volatility were related. Attanasio (1991) and Attanasio and Wadhwani (1989) show that dividend yields were a significant determinant of volatility where as Engle and Rodrigues (1989) found a positive relationship between M1 money supply, the oil price and conditional volatility.

Many theoretical models of stock prices focus on explaining stock returns with a time series model. [e.g., DeLong, Shleifer, Summers, and Waldmann, (1990), Daniel, Hirshleifer, and Subrahmanyam (1998), and Barberis, Shleifer, and Vishny (1998)]. If stock prices do not follow a random walk, then they may contain some type of temporary components.

Hence, Fama and French (1988) and Poterba and Summers (1988) have employed a model in which the natural log of stock market price consists of two components: a random walk and a temporary component. They further assume for simplicity that innovations in the two components are mutually uncorrelated. They could explain returns beyond a year by formally introducing a first-order autoregressive, AR(1), temporary component into the model of stock prices. It is noted that the long horizon returns should be explained by time aggregation of short horizon returns. Poterba and Summers (1988) extend their temporary component specification into a more general form in the estimation partly because of the above difficulties. Lo and MacKinlay (1988) point out that the construction of a single stochastic process that fits both short and long horizon returns behaviour remains a direction for further studies of stock price modeling.

The introduction of Index models came about when Sharpe (1970), realised the need for a simple approach for representing the relationships amongst securities as well as computational ease. Regardless of the introduction of these Index models by Sharpe, there was a substantial reliance on the inverse of the matrix of variances and covariances for the returns on risky securities in the derivations of the mean-variance model of portfolio selection. However, Brusser (1977) shows that such an inverse does not exist when risk-free combinations can be formed from the risky securities. The general validity of the mean-variance model is therefore challenged by the existence of opportunities for hedging. Alternative procedures, presented by Brusser (1977), can be used with or without riskless securities in conjunction with either singular or non-singular variance-covariance matrices. These re-

sults are generalised. The results are generalised and simplified statements are obtained for the frontier of mean-variance combinations and for the set of frontier portfolios.

In an earlier paper, Elton, Gruber and Padberg (1977) attempt to overcome these difficulties in generating inputs to the general portfolio model. They showed that simple decision rules, which do not involve an iterative algorithm could be derived for the selection of an optimal portfolio. The Black-Litterman (1991) methodology was developed as a result of the mean-variance analysis being extremely sensitive to the inputs for the mean and variance. This methodology alleviated the input-sensitivity problem. The methodology employs the conditional distribution theory to adjust the entire mean vector to reflect an investor's view about profiles of certain assets. A new method was presented by Qian and Gorman (2001) which extended the work of Black Litterman. Qian and Gorman's method used Bayesian inference to obtain the conditional mean vector and conditional covariance matrix when given the market equilibrium and the views of the investor.

Markowitz (1959) discovered the "asymmetrical" inefficiencies inherent in the conventional mean-variance models and suggested a semi-variance measure of asset risk that focuses only on the risks below a certain target level of return. This seemed a more attractive alternative. This measure of risk, the semi-variance, was found theoretically found to be a more robust measure of risk. (Sing and Ong, 2000). Generally, monthly returns are more or less symmetric. Since there existed no conclusive means of estimating covariance matrices, Goldman Sachs adopted methods that provided covariance matrices which were consistent with empirical regularities. Common regularities such as time-varying volatilities and correlations and fat tailed distributions were common characteristics of financial

data (Litterman and Winkelmann, 1998). Weighted returns and a decay rate through the use of a likelihood function to estimate covariances, are proposed by Litterman and Winkelmann (1998). Furthermore, they highlight alternative methods of estimating covariance matrices. These include implied volatilities, GARCH (Generalised Autoregressive conditional heteroskedastic) models, Markov chains and the notion of stationary and transitory components of stock prices.

Following up these studies, in this thesis, we explore the classical description of the Markowitz optimal portfolio selection problem, which is in nature a static and discrete-time treatment, and then pay attention to investigate the continuous counterpart – the stochastic optimal control problem of the Merton portfolio selection problem subject to a stochastic volatility diffusion process. However, instead of further pursuing a theoretical answer, which is a stochastic variation calculus problem, we turn our attention back to the parallel discrete-time development for seeking empirical evidences for identifying which stochastic volatility process family may offer optimality, which is a statistical state space modeling exercise. This thesis makes contributions in the field of portfolio optimisation and volatility modelling. It also gives rise to important empirical findings of a financial engineering nature with regards to volatility models, specifically the $GARCH(1, 1)$ case. The next section summarises the contributions of this thesis in detail.

1.2 Thesis Contributions

This thesis contains a number of contributions in modern portfolio theory. The study commences with the an empirical investigation on the superiority of the $GARCH(1, 1)$ (Au-

toregressive Conditional Heteroskedastic) model in the context of risk-return modelling, namely efficient frontier construction. *GARCH*(1, 1)'s risk-return dominance is empirically demonstrated in the single and multiple index context. Apart from its extensions within the volatility model framework, it is compared against other modelling paradigms i.e. ARMA (Autoregressive Moving Average), OLS (Ordinary Least Squares) and State Space models, where its supremacy is further established on an empirical level.

A new index model, the Improved Sharpe Model, is proposed by means of introducing a correlated variance-covariance structure for random components for single and multiple index models of Sharpe (Chapter 7). The ex post efficient portfolio investigations are carried on in terms of ARMA, OLS, State Space and *GARCH* volatility modelling on the time-series based innovation processes. Numerical evidences show us exciting signals which will produce significant impacts on investment industry. In particular the numerical evidences found in this thesis show that the Sharpe portfolio models consistently under-estimate risk in the presence of positively correlated random components, whilst the converse is observed for the negative correlation case.

In chapter 9, we back up these empirical claims through the unique application of Principal Components Analysis (PCA) within the risk-return framework. The PCA study is used to illustrate and support the claim that the Improved Sharpe model proposed by this thesis is in fact a more accurate approach than the Sharpe model in the presence of correlated residuals. Furthermore, in chapter 10, the claim and PCA study are given supplementary support through a small Monte-Carlo study that further proves the effectiveness and accuracy of the Improved Sharpe Models.

The following section outlines the objectives of this dissertation.

1.3 Objectives of the Study

The main objectives of this study can be summarised as follows:

- To demonstrate the mean-variance superiority of $GARCH(1, 1)$ compared with its volatility extensions.
- To demonstrate the superiority of the $GARCH(1, 1)$ efficient frontiers over OLS, State Space and ARMA frontiers with respect to both the Sharpe Single and Multiple index scenarios.
- To demonstrate the $GARCH(1, 1)$ frontier's superiority over its respective counterparts within the Improved Sharpe Multiple and Single index settings.
- To illustrate $GARCH(1, 1)$'s overall superiority
- To illustrate the restrictiveness of the Sharpe index models with regards to the estimation of risk-return frontiers.
- To empirically establish that Sharpe Portfolio models are under-estimating risk for portfolios exhibiting a positively correlated residual structure
- To empirically establish that Sharpe Portfolio models are over-estimating risk for portfolios exhibiting a pre-dominantly negatively correlated residual structure

- To show via PCA and Monte Carlo methods that the Improved Sharpe Portfolio Model is more accurate in illustrating risk-return relationships than the Sharpe Models under an environment presenting correlated residuals.

1.4 Limitations of the Study

This study does not consider the vast literature available that discusses the use of robust or weighted estimation. The study is also limited in the number of selection criteria used for the model selection part of the study. Selection criteria such as the Corrected Form of Akaike's Information Criterion, Hannan and Quinn Information Criterion, the Corrected Form of Hannan and Quinn (HQc) Information Criterion and many other useful model selection criterion could have been investigated in order to improve the effective selection of models. Furthermore, all empirical analysis was constrained to a financial portfolio of only 9 stocks on the Johannesburg Stock Exchange. The current portfolio does not represent the optimal diversified portfolio. In addition, the JSE-all share index and others are used as market proxies, whereas theory suggests the use of all encompassing indices that are representative of all possible securities that exist (Roll, 1977).

The study is also restricted in the variety of models that have been chosen to be fitted. Again, it would prove extremely time consuming to try and fit every possible time series model there exists in order to fit the most effective one. Numerous paths and techniques exist that one could adopt for. If one attempts to persist fitting different models, the factor

of time also becomes a crucial variable. As a result the amount of time that the study was to be completed in was limited.

The application of the methods in the study were restricted to the South African stock market during the period July 1988 to February 2005. Our analysis was also restricted to only monthly data. Only the forward selection process was implemented during model selection, whilst alternative selection procedures such as Backward elimination and All Subset/Possible Regression were not considered, thus the study was limited in terms of model selection procedures. Furthermore the study only considered $GARCH(1, 1)$'s superiority amongst a class of 4 estimation procedures. The study could have been expanded to include other types of modelling, i.e. Value-at-risk (VAR) models. Hence the study was restricted to the number of estimation techniques used in the comparisons conducted.

With regards to the exploration of the superiority of $GARCH(1, 1)$ and its extensions, the study is only conducted within the single index scenario, whilst the multiple setting was ignored. Once again given the time limit and objectives of this thesis it did would not prove practical to do so. Furthermore it provides opportunity for further research in the future.

1.5 Plan of Development of the thesis

This dissertation is comprised of ten chapters. Each chapter includes its own set of appendices and the list of appendices are added at the end of each chapter. All references however are displayed at the end of chapter eleven.

The current chapter, introduces the reader to the background and objectives of the study. It sets the scene for the reader of what is to be expected. The chapter highlights the principal focal points of the research and constructs the various hypothesis to be tested. It gives an overall summary of the research conducted.

The next five chapters, two to six, are devoted towards the modelling techniques used throughout this thesis which are to be involved in numerous comparative studies in the subsequent chapters that follow. These five introductory chapters provide the reader with theoretical background into the respective modelling frameworks and the model building process involved. A brief description of each follows.

Chapter two serves as an introduction to the fundamental concepts to financial time series modelling. Basic concepts like stationarity and log returns are introduced here which form the basic criteria required for modelling purposes. This chapter introduces the idea of time series analysis, its objectives and the fundamental concepts underlying stock theory.

Chapter three discusses *OLS* (Ordinary Least Squares), multivariate regression, model selection and introduces the various model selection criteria. This chapter establishes the basic model building process in a regression context. This chapter forms the base for chapters four and five, which expand on the models built in the empirical study of this chapter. It also introduces the model selection process that is to be adopted throughout this study. An empirical study is included in this chapter that concentrates on building multiple and single index regression models for our portfolio of stocks.

Chapter four introduces the different components and models that can be built within the time series framework. It establishes how one identifies certain time series using the

ACF and PACF and also discusses their estimation processes. It provides an examination of *ARMA* (Autoregressive Moving Average) models in general and also an empirical study that focuses on building single and multiple index time series models for our portfolio of stocks.

Chapter five presents the rivals of the time series models in this dissertation, state space modelling. Using the regression models established in chapter three, we convert these models into a *SS* (State Space) framework. We discuss the three modelling approaches within the *SS* framework and introduce the Kalman Filter, a recursive estimation tool used in estimating and building the state space model. The chapter also discusses the representation and estimation of this class of models. In the empirical study of this chapter, we develop the regression models from chapter 3 into state space representations.

Chapter six is devoted to volatility models. The respective time series and regression models built thus far are converted into a volatility framework. We also explore various extensions of the principal volatility model, *GARCH* (General Autoregressive Conditional Heteroskedastic) and make various modelling comparisons. The empirical section of this study establishes a *GARCH*(1, 1) volatility model for our portfolio of stocks in a single and multiple index context. Furthermore extensions of the *GARCH*(1, 1) model are explored in the single index context.

Up to now, chapters 3,4,5 and 6 introduce the reader to the 4 main modelling paradigms that are to be compared in a risk-return context in the chapters to follow. Chapter 7 introduces the reader to modern portfolio theory and introduces the Sharpe Single and Multiple Index models. It also presents the estimation and construction process of the

(risk-return frontier) efficient frontier of a given estimation method. This now sets up the framework for conducting comparisons of the 4 previously discussed modelling methods in a risk-return setting. The chapter goes on to highlight a very important empirical finding that illustrates that the $GARCH(1, 1)$ estimation method results in superior risk-return frontiers in both the single and multiple contexts under the Sharpe formulation. There exist many instances where the left shifting superior behaviour of $GARCH(1, 1)$ is observed. One such instance is when $GARCH(1, 1)$ is compared within the volatility modelling environ of the single index framework, i.e. we compare it with the risk-return relationship of its extensions namely $EGARCH$ (Exponential $ARCH$), $TARCH$ (Threshold $ARCH$), $PARCH$ (Power $ARCH$) and $C - GARCH$ (component $GARCH$) models.

Chapter 8, elaborates on a set of new single and multiple index portfolio models proposed by this thesis, namely the Improved Sharpe Portfolio Model. It investigates the behavior of the efficient frontier when a covariance with highly correlated residuals is introduced under the Markowitz portfolio framework. It further establishes that the Improved Sharpe Portfolio Model, proposed by Troskie and Hossain, is a more accurate and realistic outlook of the market. The empirical results in this chapter illustrate that the Sharpe Index models gives an inaccurate risk structure of an investor's financial portfolio. For portfolios exhibiting positively correlated residuals the Sharpe formulation consistently under-estimates risk whilst the converse is observed for portfolios exhibiting strong negative correlation. The consistency of this risk structure is explored under the four estimation methods namely, $GARCH(1, 1)$, $ARMA$, OLS (Ordinary Least Squares-Regression) and SS (State Space) models. This section of the thesis continues to conveniently portray the

consistent persistence of the GARCH(1,1) model's risk superiority under the various portfolio index model formulations.

Chapter 9 presents an approach, with regards to modelling indices, the principal components analysis (PCA). This chapter presents concepts of dimension reduction of data in order to simplify the view of data. We apply the Principal Components (abbreviated as PC) idea for constructing significant orthogonal components of indices in order to attain efficient frontiers that illustrate a true risk and return structure for a portfolio of stocks on the South African stock exchange under the Sharpe and Improved Sharpe formulations. The empirical evidence shows that the application of Principal Components (PC) in efficient frontier construction supports the risk-return structure estimated under the Improved Sharpe Multiple Index (ISMI) formulation. Furthermore the empirical results in this section of the study supports the claim from the previous chapter that the Sharpe Multiple Index (SMI) model gives an inaccurate risk structure of an investor's financial portfolio in the presence of positive correlation between residuals and is consistently under-estimating risk for portfolios with positively correlated residuals. This consistent risk relationship between SMI and ISMI is illustrated from a PC perspective and serves as a support for empirical claims made in earlier studies. The relationship is shown consistent under the four estimation methods: *GARCH*(1, 1), *ARMA*, *OLS* and *SS* models.

Chapter 10, provides an alternative approach, the Monte-Carlo approach, whereby stock returns are to be simulated via Monte-Carlo methods in order to illustrate that the Sharpe formulation for the multiple index framework consistently under-estimates risk for portfolios exhibiting strong positive correlation under a Monte-Carlo simulated environ

and explore the consistency of this claim under four estimation methods namely, GARCH , ARMA , OLS and SS models. A comparison is then conducted on the behaviour of the simulated portfolio under the two developed multiple index portfolio frameworks, namely the SMI and the ISMI portfolio models. An attempt will be made to show that the SMI formulation consistently under-estimates risk for portfolios exhibiting strong positive correlation under a Monte-Carlo simulated environ and explore the consistency of this claim under the four previously mentioned estimation methods.

As a consequence, the PCA and Monte-Carlo methods are used as justifications of the consistent empirical behaviour portrayed between the Sharpe and Improved Sharpe portfolio models. Furthermore these two methods also serve to elaborate on the conclusion that the Improved Sharpe models are a positive development based on the Sharpe formulations in terms of providing a more accurate risk-return structure for the investor in the presence of a correlated residual structure for financial portfolios.

Finally we conclude the study with a chapter of summary and conclusions, that explores viable aspects for future studies.

Chapter 2

Some Fundamentals of Stock Market Theory

2.1 Introduction

From Troskie (2002), a time series is a set of observations $\{x_t\}$ generated by some process, each one being recorded at a specific time t . Time series arise as recordings of processes which vary over time. A recording can either be a continuous trace or a set of discrete observations. We will concentrate on the case where observations are made at discrete equally spaced times. By appropriate choice of origin and scale we can take the observation times to be $1, 2, \dots, T$ and we can denote the observations by Y_1, Y_2, \dots, Y_T . If one repeatedly observes a given system at regular time intervals, and makes a measurement at those time intervals, it is likely that the observations are correlated. Many of the statistical modelling techniques studied relate to data which are independent or, at least, uncorrelated. Consequently, the time order of the data is important: shuffling observations will cause substantive structure to be lost from a time series. In time series analysis we exploit the correlation in the data to develop statistical modelling tools.

When fitting a parametric model to a time series, we can use a similar framework to that for modelling random samples (e.g. linear models); that is, data are explained by a model along with errors or residuals. In the case of time series, errors will themselves constitute a time series. One usually aims for the errors to be devoid of any structure, although they may be correlated. However, if one can explain the correlation in the errors,

then one ought to be left with a residual series with no correlation (or structure). Such a series is referred to as noise. Therefore, we are concerned with extracting a signal, leaving behind residual noise. (Tsay (2001))

2.2 Stationarity and Non-Stationarity

A key idea in time series is that of stationarity. Roughly speaking, a time series is stationary if its behaviour does not change over time. This means, for example, that the values always tend to vary about the same level and that their variability is constant over time. Stationary series have a rich theory and their behaviour is well understood. This means that they play a fundamental role in the study of time series.

Obviously, not all time series that we encounter are stationary. Indeed, non-stationary series tend to be the rule rather than the exception. However, many time series are related in simple ways to series which are stationary. Two important examples of this are:

1. Trend Models: The series we observe is the sum of a deterministic trend series and a stationary noise series. A simple example is the linear trend model:

$$Y_t = \beta_0 + \beta_1 t + \epsilon_t$$

Another common trend model assumes that the series is the sum of a periodic seasonal effect and stationary noise. There are many other variations.

2. Integrated models: The time series we observe satisfies

$$Y_{t+1} - Y_t = \epsilon_{t+1}$$

where ϵ_t is a stationary series. A particularly important model of this kind is the random walk. In that case, the ϵ_t values are independent shocks which affect perturb the current state Y_t by an amount ϵ_{t+1} to produce a new state Y_{t+1} .

Loosely speaking, a time series $\{X_t, t = 0, \pm 1, \dots\}$ is said to be stationary if it has statistical properties similar to those of the "time shifted" series $\{X_{t+h}, t = 0, \pm 1, \dots\}$ for each integer h . Restricting attention to those properties which depend only on the first and second order moments of $\{X_t\}$, we can make this idea precise with the following definitions.

Definition 1 *Strictly Stationarity.*

A time series $\{r_t\}$ is said to be strictly stationary if the joint distribution of

$$(r_{t_1}, \dots, r_{t_k})$$

is identical to that of

$$(r_{t_1+t}, \dots, r_{t_k+t})$$

for all t . Thus the joint distribution of $(r_{t_1}, \dots, r_{t_k})$ is invariant under time shift.

Definition 2 *Weak Stationarity.*

A time series $\{r_t\}$ is said to be weakly stationary if both the mean of r_t ,

$$E(r_t) = \mu \tag{1}$$

and the covariance between r_t and r_{t-l}

$$\text{cov}(r_t, r_{t-l}) = \gamma_l \tag{2}$$

is time invariant and finite. In practice, weak stationarity implies that if we observe a time series $\{r_t, t = 1, \dots, T\}$ then the time plot of the data would show that the T values fluctuate with constant variation around a constant level. Figure 2.1 below illustrates the time plot for the stock Anglos, which clearly shows that there is a multiplicative trend, hence the prices are increasing in an exponential manner. It also confirms that there is no seasonality present, otherwise we would become millionaires.

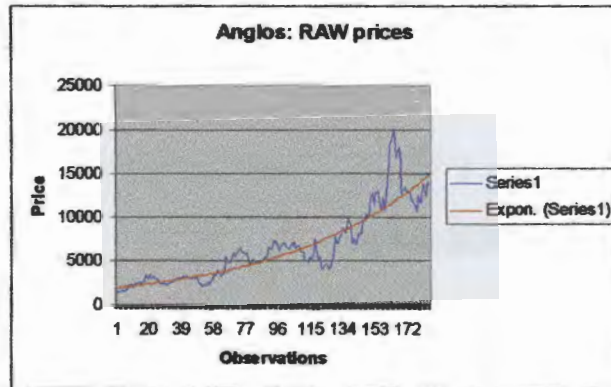


Figure 2.1: Time Plot of Raw prices for ANGLOS

Figures 2.2 and 2.3 below depict the time plots of the log returns of Sappi and Anglos respectively there is a clear indication of mean stationarity, but can we say the same of the variances and covariances. Referring back to figure 2.1 above it is clear that the time series of raw prices for Anglos contains a multiplicative trend, thus the values P_t , the raw price of the stock, will not have a constant mean or variance thus it is non-stationary. Implicitly in the condition of weak stationarity, we assume that the first two moments of r_t are finite. If r_t is strictly stationary and its first two moments are finite then r_t is also weakly stationary. The converse is not true in general. However, if the time series is normally distributed, then weak stationarity is equivalent to strict stationarity. In this study we will be mainly concerned with weakly stationary series.

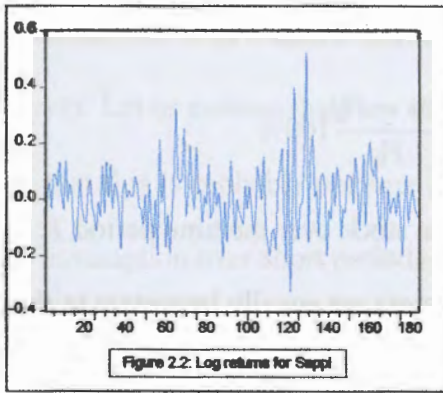


Figure 2.2: Log returns for Suppl

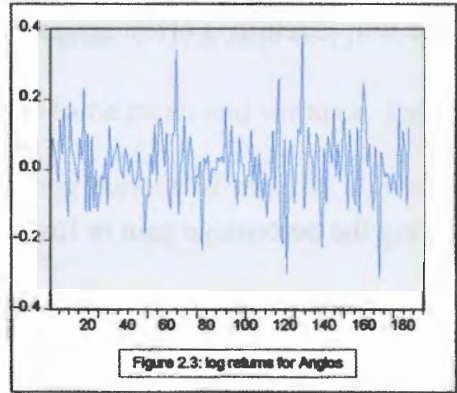


Figure 2.3: log returns for Anglos

The covariance $cov(r_t, r_{t-l}) = \gamma_l$ is called the lag- l autocovariance of r_t . It has two

important properties

$$(a) \gamma_0 = var(r_t)$$

$$(b) \gamma_l = \gamma_{-l}.$$

In the finance literature, it is common to assume that the asset or stock return series is weakly stationary. This assumption is often checked empirically. (Troskie (2002))

2.3 Some Financial Stock Theory: Fundamental Concepts

This thesis attempts to apply principally two different modelling approaches to the log returns of four JSE listed stocks. Before commencing the modelling aspect, an introduction to stock theory is preliminarily required. Let P_T be the price of a stock or asset at some time T (past or future). Let P_0 be the price at time T_0 - which could be the beginning price at the time horizon of interest. The portfolio manager or investor is interested in the price change $P_T - P_0$, or more particularly, the relative price change called the simple return

$$R_T = \frac{P_T - P_0}{P_0}.$$

The simple return is often expressed as a percentage

$$\frac{P_T - P_0}{P_0} 100\%$$

giving the percentage gain or loss of the stock over the time period T . Gains (positive percentages) and losses (negative percentages) are equally important to the portfolio manager or investor. In this study whilst doing the calculations it was preferred to work with the actual simple returns rather than the percentages, but the percentage simple return is often the one quoted.

The price P_T is considered to be a random variable having some distribution. The following theory has been developed by famous financial economists (econometricians) and statisticians (some are Nobel price winners).

Suppose there are n time periods in the time difference zone $T - T_0$. Suppose further that these time periods are equally spaced in time (i.e. hours, days, weeks, months, quarters, years etc.) for this study, monthly returns were used. Let us denote these prices by P_t , $t = 1, \dots, n$ and let the simple returns be

$$R_t = \frac{P_t - P_{t-1}}{P_{t-1}}, \quad t = 1, \dots, n. \quad (1)$$

Note that $P_T = P_n$. Clearly any P_t or R_t is also a random variable with some mean (expected value) and variance. From the compound interest formula in Finance

$$(1 + R_T) = (1 + R_1)(1 + R_2) \dots (1 + R_n) \quad (2)$$

so that

$$\log(1 + R_T) = \sum_{t=1}^n \log(1 + R_t). \quad (3)$$

A crucial assumption is now made. It is tested regularly. The $R_t, t = 1, \dots, n$ are random variables (rvs). Let us assume that they have the same mean and variance. Past experience has shown that this is probably not true over long periods of time. It is, however a very reasonable assumption over short periods of time.

We also make the assumption that the R_t are independently distributed. If they are not, then they are dependent and hence is a time series. Statisticians know how to use time series to forecast. So, if stock price returns R_t are dependent then statisticians will be enormously rich. We know that they are poor. So stock price returns cannot be dependent, i.e. a time series.

Thus our final crucial assumption is that stock price returns R_t are independently and identically distributed (i.i.d.) with the same mean and variance. From the Central Limit Theorem then follows that

$$\log(1 + R_T) = \sum_{t=1}^n \log(1 + R_t), \quad (4)$$

which is the sum of n i.i.d. random variables, follows a Normal distribution with some mean and variance. Now

$$\begin{aligned} \log(1 + R_T) &= \log\left(1 + \frac{P_n - P_0}{P_0}\right), \text{ or} \\ \log(1 + R_T) &= \log\left(1 + \frac{P_T - P_0}{P_0}\right) \\ &= \log(P_T/P_0) \\ &= \log P_T - \log P_0. \\ &= p_T - p_0 \\ &= r_T \end{aligned}$$

where

$$p_T = \log P_T. \quad (6)$$

Thus the random variable

$$r_T = \log P_T - \log P_0 = p_T - p_0 \quad (7)$$

is distributed (for large n) as a Normal distribution with mean $E(r_T) = \mu$ and variance $\text{var}(r_T) = \sigma^2$. That is $r_T \simeq N(\mu, \sigma^2)$.

Depending on the time spacings t the number n can be very large. Thus

$$\log P_T = \log P_0 + r_T \text{ or } P_T = P_0 e^{r_T}. \quad (8)$$

If $X \sim N(\mu, \sigma^2)$ then the moment generating function of X is

$$\begin{aligned} M_X(t) &= E(e^{tX}) \\ &= \exp\left(t\mu + \frac{1}{2}t^2\sigma^2\right). \end{aligned}$$

From

$$P_T = P_0 e^{r_T} \quad (9)$$

so that

$$E(P_T) = P_0 e^{\mu + \sigma^2/2} \quad (10)$$

which is a fundamental equation in Finance. The random variable

$$r_T = \log P_T - \log P_0 = p_T - p_0$$

is often referred to as the log return and the statement is made that log returns are normally distributed. Alternatively, the statement is made that stock prices P_T are log-normally distributed. In future when we say returns we will mean log returns. In practice, the simple

returns

$$R_T = \frac{P_T - P_0}{P_0}$$

are used rather than the log returns $r_T = \log P_T - \log P_0$. In most calculations it will not make any difference whether we use simple returns or log returns, but in this theses we work with log returns, unless specified otherwise. In general we would expect more positive returns (positive skewness) than negative returns (negative skewness). Otherwise nobody will invest in stocks. Taking logs pulls in the right tail more to the centre and pushes the left tail out. It makes the distribution therefore more symmetric, and hence working with log returns is more desirable for estimation purposes. For symmetric distributions the sample mean is often a good estimate (the best for the normal case).

The log return $r_T = \log P_T - \log P_0$ is for large n approximately normally distributed $N(\mu, \sigma^2)$. This means that the time span $T - T_0$ must be large enough for the Central Limit Theorem to hold. We quote from Fama "The Foundations of Finance." The frequency distributions, the comparisons of average relative frequencies with normal probabilities and the studentized ranges, all lead to the conclusion that distributions of monthly returns are closer to normal than distributions of daily returns. This is inconsistent with the hypothesis that return distributions are non-normal symmetric stable, which implies that distributions of daily and monthly returns should have about the same degree of leptokurtosis. Moreover, although the evidence also suggests that distributions of monthly returns are slightly leptokurtic relative to normal distributions, let us tentatively accept the normal model as a working approximation for monthly returns.

The assumptions that distributions of returns on stocks and portfolios are normally distributed is used to develop a model for portfolio decisions by investors. The usefulness of the portfolio model depends not on whether the normality assumption which underlies it, is an exact description of the world (we know it is not), but on whether the model yields useful insights into the essential ingredients of rational portfolio decision. If the model does well on this score, we can live with the small observed departures from normality in monthly returns, all least until better models come along". (Troskie 2002)

2.3.1 Distributional Properties of Stock Returns.

It has been demonstrated that log returns r_t have desirable statistical properties. We now study these properties further. Consider a collection of p stocks or assets observed, or held, for T time periods, say $t = 1, 2, \dots, T$. For each stock i , let r_{it} be the log return at time t . Thus if

$$\{P_{it}, i = 1, \dots, p, t = 1, \dots, T\} \quad (10)$$

is the price of stock i at time t then

$$\{p_{it} = \log P_{it}, i = 1, \dots, p, t = 1, \dots, T\} \quad (11)$$

is the log price and

$$r_{it} = p_{it} - p_{it-1}, i = 1, \dots, p, t = 1, \dots, T \quad (12)$$

is the log return; the main variable under study. Of interest also, to us, is the simple return

$$\{R_{it} = \frac{P_{it} - P_{it-1}}{P_{it-1}}, i = 1, \dots, p, t = 1, \dots, T\} \quad (13)$$

and the excess simple and log returns

$$\{XR_{it}, = R_{it} - R_{fit}, i = 1, \dots, p, t = 1, \dots, T\} \quad (14)$$

$$\{xr_{it} = r_{it} - r_{ift}, i = 1, \dots, p, t = 1, \dots, T\}$$

with R_{fit} and r_{ift} the risk free rate *returns*, respectively. Let the distribution function of the log returns

$$\{r_{it} \mid i = 1, \dots, p, t = 1, \dots, T\} \quad (15)$$

be

$$F_r[\{r_{ij}\}/\mathbf{Y}, \boldsymbol{\theta}] \quad (16)$$

where \mathbf{Y} is a state vector consisting of variables that describes the behaviour of stock log returns and $\boldsymbol{\theta}$ is a vector of parameters. The state vector is assumed given and the main purpose is to estimate the parameters $\boldsymbol{\theta}$, and draw inference about behaviour of $\{r_{ij}\}$ given some past data. Of interest is the behaviour of

$$\mathbf{r}_t = \begin{pmatrix} r_{1t} \\ \vdots \\ r_{pt} \end{pmatrix} \quad (17)$$

for a single time period t , like in modern portfolio theory, or in

$$\mathbf{r}_i = \begin{pmatrix} r_{i1} \\ \vdots \\ r_{iT} \end{pmatrix} \quad (18)$$

for a single stock i , as in financial time series. In the latter case we usually denote the time series by

$$\mathbf{r}'_i = (r_{i1}, \dots, r_{iT}) \quad (19)$$

or even more conveniently drop the subscript i . In this study we will consider both theories.

For the time series (19) dropping the subscript i let the joint distribution of the time series

(r_1, \dots, r_T) be

$$\begin{aligned} F(r_1, \dots, r_T/\theta) &= F(r_1/\theta)F(r_2/r_1, \theta) \dots F(r_T/r_{T-1}, \dots, r_1, \theta) \\ &= F(r_1/\theta) \prod_{t=2}^T F(r_t/r_{t-1}, \dots, r_1, \theta). \end{aligned} \quad (20)$$

Thus of main interest is the behaviour of the conditional distribution

$$F(r_t/r_{t-1}, \dots, r_1, \theta). \quad (21)$$

and in particular how the conditional distribution $F(r_t/r_{t-1}, \dots, r_1, \theta)$ evolves over time.

Since, on the JSE we only have low frequency data we may assume that the distribution is continuous with probability density function

$$f(r_1, \dots, r_T/\theta) = f(r_1/\theta) \prod_{t=2}^T f(r_t/r_{t-1}, \dots, r_1, \theta). \quad (22)$$

For high frequency log returns, as for the New York Stock Exchange (NYSE), discreteness becomes an issue. In the next section we give a brief summary of the properties of the univariate and multivariate normal distributions. These two distributions play a major role in the study that follows.

2.4 The Univariate Normal Distribution. Tests for Normality.

If (x_1, \dots, x_T) is a sample from a random variable X which has a normal distribution, say

$$X \sim N(\mu, \sigma^2) \quad (1)$$

then the sample mean (m.l.e of μ) is

$$\bar{x} = \frac{1}{T} \sum_{t=1}^T x_i, \quad (2)$$

, the sample variance is

$$s^2 = \frac{1}{T-1} \sum_{t=1}^T (x_t - \bar{x})^2. \quad (3)$$

The sample skewness is

$$b_1 = \frac{\frac{1}{T-1} \sum_{t=1}^T (x_t - \bar{x})^3}{s^3}, \quad (4)$$

and the sample kurtosis is

$$b_2 = \frac{\frac{1}{T-1} \sum_{t=1}^T (x_t - \bar{x})^4}{s^4}. \quad (5)$$

Under the assumption of normality of X , b_1 and b_2 are asymptotically normally distributed with

$$b_1 \approx N(0, 6/T) \quad (6)$$

and

$$b_2 \approx N(3, 24/T). \quad (7)$$

Thus

$$Z_1 = \frac{b_1}{\sqrt{6/T}} \text{ and } Z_2 = \frac{b_2 - 3}{\sqrt{24/T}} \quad (8)$$

are asymptotically $N(0, 1)$ distributed. These statistics can therefore be used to test for normality when T is large. The statistic

$$b_2 - 3$$

is often used and called the excess kurtosis. An omnibus test for normality is

$$\chi^2 = Z_1^2 + Z_2^2 \sim \chi^2 \quad (9)$$

with 2 degrees of freedom. Since for stock price data T is usually large the above tests can be used. For small T more exact tests are available. In the following section these test statistics and some graphical displays are given on the stock price data of the JSE-

, the sample variance is

$$s^2 = \frac{1}{T-1} \sum_{t=1}^T (x_t - \bar{x})^2. \quad (3)$$

The sample skewness is

$$b_1 = \frac{\frac{1}{T-1} \sum_{t=1}^T (x_t - \bar{x})^3}{s^3}, \quad (4)$$

and the sample kurtosis is

$$b_2 = \frac{\frac{1}{T-1} \sum_{t=1}^T (x_t - \bar{x})^4}{s^4}. \quad (5)$$

Under the assumption of normality of X , b_1 and b_2 are asymptotically normally distributed with

$$b_1 \approx N(0, 6/T) \quad (6)$$

and

$$b_2 \approx N(3, 24/T). \quad (7)$$

Thus

$$Z_1 = \frac{b_1}{\sqrt{6/T}} \text{ and } Z_2 = \frac{b_2 - 3}{\sqrt{24/T}} \quad (8)$$

are asymptotically $N(0, 1)$ distributed. These statistics can therefore be used to test for normality when T is large. The statistic

$$b_2 - 3$$

is often used and called the excess kurtosis. An omnibus test for normality is

$$\chi^2 = Z_1^2 + Z_2^2 \sim \chi^2 \quad (9)$$

with 2 degrees of freedom. Since for stock price data T is usually large the above tests can be used. For small T more exact tests are available. In the following section these test statistics and some graphical displays are given on the stock price data of the JSE-

, the sample variance is

$$s^2 = \frac{1}{T-1} \sum_{t=1}^T (x_t - \bar{x})^2. \quad (3)$$

The sample skewness is

$$b_1 = \frac{\frac{1}{T-1} \sum_{t=1}^T (x_t - \bar{x})^3}{s^3}, \quad (4)$$

and the sample kurtosis is

$$b_2 = \frac{\frac{1}{T-1} \sum_{t=1}^T (x_t - \bar{x})^4}{s^4}. \quad (5)$$

Under the assumption of normality of X , b_1 and b_2 are asymptotically normally distributed with

$$b_1 \approx N(0, 6/T) \quad (6)$$

and

$$b_2 \approx N(3, 24/T). \quad (7)$$

Thus

$$Z_1 = \frac{b_1}{\sqrt{6/T}} \text{ and } Z_2 = \frac{b_2 - 3}{\sqrt{24/T}} \quad (8)$$

are asymptotically $N(0, 1)$ distributed. These statistics can therefore be used to test for normality when T is large. The statistic

$$b_2 - 3$$

is often used and called the excess kurtosis. An omnibus test for normality is

$$\chi^2 = Z_1^2 + Z_2^2 \sim \chi^2 \quad (9)$$

with 2 degrees of freedom. Since for stock price data T is usually large the above tests can be used. For small T more exact tests are available. In the following section these test statistics and some graphical displays are given on the stock price data of the JSE-

Over Index, and Anglos, a share listed on the JSE (Johannesburg Stock Exchange). Both these financial time series play an important role in the theoretical and practical results that follow. (Troskie (2002))

2.4.1 The Multivariate Normal Distribution.

Consider a vector

$$X = \begin{pmatrix} X_1 \\ \vdots \\ X_p \end{pmatrix}$$

of random variables X_i (stock returns) with expected value (mean)

$$E(X) = \mu$$

and covariance matrix

$$\Sigma = E(X - \mu)(X - \mu)'$$

If X has a density of the form

$$f(x_1, \dots, x_p) = \frac{1}{(2\pi)^{p/2} |\Sigma|^{1/2}} \exp\left[-\frac{1}{2}(x - \mu)' \Sigma^{-1}(x - \mu)\right] \quad (1)$$

then X has a multivariate normal distribution

$$X \sim N(\mu, \Sigma).$$

The multivariate normal distribution has the following important properties. If $Y = CX$ for any matrix C then $Y \sim N(C\mu, C\Sigma C')$, that is, linear combinations of multivariate normal variates are again multivariate normally distributed.

If X is partitioned

$$X = \begin{pmatrix} X^{(1)} \\ X^{(2)} \end{pmatrix}$$

into two sets then

$$X^{(1)} \sim N(\mu^{(1)}, \Sigma_{11}) \text{ and}$$

$$X^{(2)} \sim N(\mu^{(2)}, \Sigma_{22})$$

where

$$\mu = \begin{pmatrix} \mu^{(1)} \\ \mu^{(2)} \end{pmatrix} \text{ and } \Sigma = \begin{pmatrix} \Sigma_{11} & \Sigma_{12} \\ \Sigma_{21} & \Sigma_{22} \end{pmatrix} \quad (2)$$

and partitioned conformably. Thus marginal distributions of multivariate normal distributions are again multivariate normally distributed. In particular $X_i \sim N(\mu_i, \sigma_{ii})$. The conditional distribution of $X^{(1)}$ given $X^{(2)} = x^{(2)}$ is also multivariate normal with distribution

$$X^{(1)}/x^{(2)} \sim N(\mu^{(1)} + \Sigma_{12}\Sigma_{22}^{-1}(x^{(2)} - \mu^{(2)}), \Sigma_{11.2}) \quad (3)$$

with

$$\Sigma_{11.2} = \Sigma_{11} - \Sigma_{12}\Sigma_{22}^{-1}\Sigma_{21} \quad (4)$$

where Σ_{ij} are given in equation (3) above. The conditional mean

$$E(X^{(1)}/x^{(2)}) = \mu^{(1)} + \Sigma_{12}\Sigma_{22}^{-1}(x^{(2)} - \mu^{(2)}) \quad (5)$$

and covariance matrix

$$\text{cov}(X^{(1)}/x^{(2)}) = \Sigma_{11.2} \quad (6)$$

play an important part in portfolio analysis. For example, if $p = 2$, and any return of a stock $R = X_1$, with the return of a market proxy, say $M = X_2$ then from the assumption

of bivariate normality

$$\begin{aligned}
 E(R/M) &= \mu_1 + \frac{\sigma_{12}}{\sigma_{22}}(M - \mu_2) \\
 &= \mu_1 - \frac{\sigma_{12}}{\sigma_{22}}\mu_2 + \frac{\sigma_{12}}{\sigma_{22}}M \\
 &= \alpha + \beta M
 \end{aligned}$$

or more conveniently written as

$$R = \alpha + \beta M + e, \quad E(e) = 0, \quad E(e^2) = \sigma^2 \quad (7)$$

which is the well known Market model or Sharp Single Index model. If any stock return $R = X_1$ and market indices $M_1 = X_2, M_2 = X_3, \dots, M_k = X_p$ then it is easy to show that from (5)

$$R = \beta_0 + \beta_1 M_1 + \dots + \beta_k M_k + e \quad (8)$$

with $E(e) = 0, \text{var}(e) = E(e^2) = \sigma^2$. This is the well known Sharpe Multi - Index model. Furthermore if $X^{(1)} = R^{(1)}$ a vector of stock returns and $M^{(2)} = X^{(2)}$, a vector of market proxies and/or other variables, then from (5)

$$E(R^{(1)}/M) = \mu^{(1)} - \Sigma_{12}\Sigma_{22}^{-1}\mu^{(2)} + \Sigma_{12}\Sigma_{22}^{-1}M^{(2)} \quad \text{or} \quad (9)$$

$$R^{(1)} = \Gamma_0 + \Gamma_1 M^{(2)} + \Xi$$

with

$$E(\Xi) = \mathbf{0}, \quad \text{and} \quad E(\Xi\Xi') = \Psi$$

which is the vector market model and includes the Troskie Innovation model. The model is also a special case of the Kalman Filter (i.e. including a time subscript t in $R_t^{(1)}$ and $M_t^{(2)}$ and the parameters). All the above models play a fundamental role in portfolio theory.

2.4.2 The Conditional Multivariate Normal Distribution

Let \mathbf{X} be distributed like a multivariate normal distribution $\mathbf{X} \sim N(\boldsymbol{\mu}, \boldsymbol{\Sigma})$ and let \mathbf{X} be partitioned as

$$\mathbf{X} = \begin{pmatrix} \mathbf{X}^{(1)} \\ \mathbf{X}^{(2)} \end{pmatrix}$$

into q and $r = p - q$ components respectively. Partition $\boldsymbol{\mu}$ and $\boldsymbol{\Sigma}$ conformably, that is

$$\boldsymbol{\mu} = \begin{pmatrix} \boldsymbol{\mu}^{(1)} \\ \boldsymbol{\mu}^{(2)} \end{pmatrix} \text{ and } \boldsymbol{\Sigma} = \begin{pmatrix} \boldsymbol{\Sigma}_{11} & \boldsymbol{\Sigma}_{12} \\ \boldsymbol{\Sigma}_{21} & \boldsymbol{\Sigma}_{22} \end{pmatrix}.$$

Consider the non-singular linear transformation

$$\mathbf{Y}^{(1)} = \mathbf{X}^{(1)} + \mathbf{M}\mathbf{X}^{(2)} \quad (1)$$

$$\mathbf{Y}^{(2)} = \mathbf{X}^{(2)}$$

choosing \mathbf{M} so that the components of $\mathbf{Y}^{(1)}$ are uncorrelated with the components of $\mathbf{Y}^{(2)}$.

The matrix \mathbf{M} must satisfy the equation

$$\begin{aligned} 0 &= E(\mathbf{Y}^{(1)} - E\mathbf{Y}^{(1)})(\mathbf{Y}^{(2)} - E\mathbf{Y}^{(2)})' \\ &= E(\mathbf{X}^{(1)} + \mathbf{M}\mathbf{X}^{(2)} - E\mathbf{X}^{(1)} - \mathbf{M}E\mathbf{X}^{(2)})(\mathbf{X}^{(2)} - E\mathbf{X}^{(2)})' \\ &= E(\mathbf{X}^{(1)} - E\mathbf{X}^{(1)} + \mathbf{M}(\mathbf{X}^{(2)} - E\mathbf{X}^{(2)}))(\mathbf{X}^{(2)} - E\mathbf{X}^{(2)})' \\ &= \boldsymbol{\Sigma}_{12} + \mathbf{M}\boldsymbol{\Sigma}_{22}. \end{aligned}$$

Thus

$$\mathbf{M} = -\boldsymbol{\Sigma}_{12}\boldsymbol{\Sigma}_{22}^{-1} \text{ and}$$

$$\mathbf{Y}^{(1)} = \mathbf{X}^{(1)} - \boldsymbol{\Sigma}_{12}\boldsymbol{\Sigma}_{22}^{-1}\mathbf{X}^{(2)}.$$

The vector

$$\begin{aligned} \mathbf{Y} &= \begin{pmatrix} \mathbf{Y}^{(1)} \\ \mathbf{Y}^{(2)} \end{pmatrix} \\ &= \begin{pmatrix} \mathbf{I} & -\Sigma_{12}\Sigma_{22}^{-1} \\ \mathbf{0} & \mathbf{I} \end{pmatrix} \begin{pmatrix} \mathbf{X}^{(1)} \\ \mathbf{X}^{(2)} \end{pmatrix} \\ &= \mathbf{C}\mathbf{X} \end{aligned}$$

is a non-singular linear transformation of \mathbf{X} and therefor has distribution

$$\mathbf{Y} \sim N(\mathbf{C}\boldsymbol{\mu}, \mathbf{C}\Sigma\mathbf{C}')$$

with

$$\begin{aligned} \mathbf{C}\boldsymbol{\mu} &= \begin{pmatrix} \mathbf{I} & -\Sigma_{12}\Sigma_{22}^{-1} \\ \mathbf{0} & \mathbf{I} \end{pmatrix} \begin{pmatrix} \boldsymbol{\mu}^{(1)} \\ \boldsymbol{\mu}^{(2)} \end{pmatrix} \\ &= \begin{pmatrix} \boldsymbol{\mu}^{(1)} - \Sigma_{12}\Sigma_{22}^{-1}\boldsymbol{\mu}^{(2)} \\ \boldsymbol{\mu}^{(2)} \end{pmatrix} \\ &= \boldsymbol{\nu} \end{aligned}$$

and

$$\begin{aligned} \mathbf{C}\Sigma\mathbf{C}' &= \begin{pmatrix} \mathbf{I} & -\Sigma_{12}\Sigma_{22}^{-1} \\ \mathbf{0} & \mathbf{I} \end{pmatrix} \begin{pmatrix} \Sigma_{11} & \Sigma_{12} \\ \Sigma_{21} & \Sigma_{22} \end{pmatrix} \begin{pmatrix} \mathbf{I} & \mathbf{0} \\ -\Sigma_{22}^{-1}\Sigma_{21} & \mathbf{I} \end{pmatrix} \\ &= \begin{pmatrix} \Sigma_{11} - \Sigma_{12}\Sigma_{22}^{-1}\Sigma_{21} & \mathbf{0} \\ \mathbf{0} & \Sigma_{22} \end{pmatrix} \\ &= \begin{pmatrix} \Sigma_{11,2} & \mathbf{0} \\ \mathbf{0} & \Sigma_{22} \end{pmatrix}, \text{ with } \Sigma_{11,2} = \Sigma_{11} - \Sigma_{12}\Sigma_{22}^{-1}\Sigma_{21} \end{aligned}$$

Thus $\mathbf{Y}^{(1)}$ and $\mathbf{Y}^{(2)}$ are independently distributed with marginal distributions

$$\mathbf{Y}^{(1)} \sim N(\boldsymbol{\mu}^{(1)} - \Sigma_{12}\Sigma_{22}^{-1}\boldsymbol{\mu}^{(2)}, \Sigma_{11} - \Sigma_{12}\Sigma_{22}^{-1}\Sigma_{21} = \Sigma_{11,2})$$

$$\mathbf{Y}^{(2)} \sim N(\boldsymbol{\mu}^{(2)}, \Sigma_{22}).$$

Using the same transformation of (1) to transform back from \mathbf{Y} to \mathbf{X} the Jacobian of transformation being one The joint density of

$$\begin{aligned}
 f(\mathbf{x}^{(1)}, \mathbf{x}^{(2)}) &= n(\mathbf{y}^{(1)}/\boldsymbol{\mu}^{(1)} - \boldsymbol{\Sigma}_{12}\boldsymbol{\Sigma}_{22}^{-1}\boldsymbol{\mu}^{(2)}, \boldsymbol{\Sigma}_{11.2}) \cdot n(\mathbf{y}^{(2)}/\boldsymbol{\mu}^{(2)}, \boldsymbol{\Sigma}_{22}) \cdot |J| \\
 &= (2\pi)^{-q/2} |\boldsymbol{\Sigma}_{11.2}|^{-1/2} \exp\{-1/2(\mathbf{x}^{(1)} - \boldsymbol{\Sigma}_{12}\boldsymbol{\Sigma}_{22}^{-1}\mathbf{x}^{(2)} - (\boldsymbol{\mu}^{(1)} - \boldsymbol{\Sigma}_{12}\boldsymbol{\Sigma}_{22}^{-1}\boldsymbol{\mu}^{(2)}))' \boldsymbol{\Sigma}_{11.2}^{-1} \\
 &\quad \cdot (\mathbf{x}^{(2)} - \boldsymbol{\mu}^{(2)})'\boldsymbol{\Sigma}_{22}^{-1} \\
 &= n(\mathbf{x}^{(1)} - (\boldsymbol{\mu}^{(1)} + \boldsymbol{\Sigma}_{12}\boldsymbol{\Sigma}_{22}^{-1}(\mathbf{x}^{(2)} - \boldsymbol{\mu}^{(2)})), \boldsymbol{\Sigma}_{11.2}) \cdot n(\mathbf{x}^{(2)}/\boldsymbol{\mu}^{(2)}, \boldsymbol{\Sigma}_{22}) \\
 &= n(\mathbf{x}/\boldsymbol{\mu}, \boldsymbol{\Sigma}).
 \end{aligned}$$

The conditional distribution of $\mathbf{X}^{(1)}$ given $\mathbf{X}^{(2)} = \mathbf{x}^{(2)}$ is then

$$\begin{aligned}
 f(\mathbf{x}^{(1)}/\mathbf{x}^{(2)}) &= \frac{n(\mathbf{x}/\boldsymbol{\mu}, \boldsymbol{\Sigma})}{n(\mathbf{x}^{(2)}/\boldsymbol{\mu}^{(2)}, \boldsymbol{\Sigma}_{22})} \\
 &= n(\mathbf{x}^{(1)} - (\boldsymbol{\mu}^{(1)} + \boldsymbol{\Sigma}_{12}\boldsymbol{\Sigma}_{22}^{-1}(\mathbf{x}^{(2)} - \boldsymbol{\mu}^{(2)})), \boldsymbol{\Sigma}_{11.2})
 \end{aligned}$$

which is a multivariate normal density with conditional mean

$$E(\mathbf{X}^{(1)}/\mathbf{x}^{(2)}) = \boldsymbol{\mu}^{(1)} + \boldsymbol{\Sigma}_{12}\boldsymbol{\Sigma}_{22}^{-1}(\mathbf{x}^{(2)} - \boldsymbol{\mu}^{(2)})$$

and conditional covariance matrix

$$\text{cov}(\mathbf{X}^{(1)}/\mathbf{x}^{(2)}) = \boldsymbol{\Sigma}_{11.2} = \boldsymbol{\Sigma}_{11} - \boldsymbol{\Sigma}_{12}\boldsymbol{\Sigma}_{22}^{-1}\boldsymbol{\Sigma}_{21}.$$

Note that the conditional mean $E(\mathbf{X}^{(1)}/\mathbf{x}^{(2)})$ depends on $\mathbf{x}^{(2)}$ but the conditional covariance matrix $\text{cov}(\mathbf{X}^{(1)}/\mathbf{x}^{(2)})$ does not depend on $\mathbf{x}^{(2)}$. This extremely important result plays a vital role in Modern Portfolio Analysis. The matrix $\boldsymbol{\Sigma}_{12}\boldsymbol{\Sigma}_{22}^{-1}$ is the matrix of regression

coefficients. The elements of

$$\begin{aligned}\Sigma_{11.2} &= \Sigma_{11} - \Sigma_{12}\Sigma_{22}^{-1}\Sigma_{21} \\ &= [\sigma_{ij.q+1,\dots,p}], \quad i, j = 1, \dots, q\end{aligned}$$

are the conditional variances and covariances also referred to as partial variances and covariances. The conditional or partial correlations are then

$$[\rho_{ij.q=1,\dots,p}] = \left[\frac{\sigma_{ij.q+1,\dots,p}}{\sqrt{\sigma_{ii.q+1,\dots,p}\sigma_{jj.q+1,\dots,p}}} \right], \quad i, j = 1, \dots, q.$$

If the correlation matrix

$$\mathbf{P} = [\rho_{ij}], \quad i, j = 1, \dots, p$$

is partitioned conformably to Σ , as

$$\mathbf{P} = \begin{pmatrix} \mathbf{P}_{11} & \mathbf{P}_{12} \\ \mathbf{P}_{21} & \mathbf{P}_{22} \end{pmatrix}$$

yields the matrix

$$\begin{aligned}\mathbf{P}_{11.2} &= \mathbf{P}_{11} - \mathbf{P}_{12}\mathbf{P}_{22}^{-1}\mathbf{P}_{21} \\ &= [\rho_{ij.q+1,\dots,p}], \quad i, j = 1, \dots, q.\end{aligned}$$

is called the partial correlation matrix. (Troskie (2002))

2.5 The Correlation and Autocorrelation Functions.

The correlation between two random variables X and Y is defined as

$$\begin{aligned}\rho_{XY} &= \frac{\text{cov}(X, Y)}{\sqrt{\text{var}(X)\text{var}(Y)}} \\ &= \frac{E(X - E(X))(Y - E(Y))}{\sqrt{E(X - E(X))^2 E(Y - E(Y))^2}}\end{aligned}$$

and if a sample $(x_i, y_i, i = 1, \dots, T)$ is available then the consistent estimator of ρ_{XY} is

$$\hat{\rho}_{XY} = \frac{\sum_{i=1}^T (x_i - \bar{x})(y_i - \bar{y})}{\{\sum_{i=1}^T (x_i - \bar{x})^2 \sum_{i=1}^T (y_i - \bar{y})^2\}^{1/2}}. \quad (1)$$

Consider now a weakly stationary return series r_t . The linear dependence between r_t and r_{t-l} is now of interest. The correlation coefficient between r_t and r_{t-l} is defined as

$$\begin{aligned} \rho_l &= \frac{\text{cov}(r_t, r_{t-l})}{\sqrt{\text{var}(r_t)\text{var}(r_{t-l})}} \\ &= \frac{\text{cov}(r_t, r_{t-l})}{\text{var}(r_t)} \\ &= \frac{\gamma_l}{\gamma_0} \end{aligned} \quad (2)$$

and is called the lag- l autocorrelation. The Autocorrelation Function (ACF) provides a useful measure of the degree of dependence between the values of a time series at different times and for this reason play an important role when we consider the prediction of future values of the series in terms of the past and present values. From the definition

$$\rho_0 = 1$$

$$\rho_l = \rho_{-l}$$

$$-1 \leq \rho_l \leq 1.$$

A weakly stationary series is not correlated if and only if $\rho_l = 0$ for all $l > 0$. For a given sample of returns $\{r_t\}_{t=1}^T$ the lag-1 sample autocorrelation of r_t is

$$\hat{\rho}_1 = \frac{\sum_{t=2}^T (r_t - \bar{r})(r_{t-1} - \bar{r})}{\sum_{t=1}^T (r_t - \bar{r})^2}. \quad (3)$$

Under some general conditions $\hat{\rho}_1$ is a consistent estimator of ρ_1 . If $\{r_t\}$ is an independent and identically distributed (*i.i.d.*) sequence and $E(r_t)^2 < \infty$ then

$$\hat{\rho}_1 \simeq N(0, 1/T) \quad (4)$$

which can be used to test the hypotheses

$$H_0 : \rho_1 = 0 \text{ vs } H_1 : \rho_1 \neq 0.$$

The lag- l sample autocorrelation is defined as

$$\hat{\rho}_l = \frac{\sum_{t=l+1}^T (r_t - \bar{r})(r_{t-l} - \bar{r})}{\sum_{t=1}^T (r_t - \bar{r})^2}, \quad 0 \leq l < T - 1. \quad (5)$$

Under the same conditions stated above

$$\hat{\rho}_l \simeq N(0, 1/T)$$

for any fixed positive integer l . More generally, if $\{r_t\}_{t=1}^T$ is a weakly stationary time series satisfying

$$r_t = \mu + \sum_{i=0}^q \psi_i e_{t-i}, \text{ where } \psi_0 = 1$$

and $\{e_t\}$ is a Gaussian white noise series (i.e. $e_t \sim N(0, 1)$) then

$$\hat{\rho}_l \approx N(0, (1 + 2 \sum_{i=0}^q \rho_i^2)/T) \text{ for } l > q. \quad (6)$$

The formula (6) is referred to as Bartlett's formula. In finite samples $\hat{\rho}_l$ is a biased estimator of ρ_l . The bias is of the order $1/T$, which can be substantial if T is small. In most financial calculations T is relatively large so that the bias is not serious.

2.5.1 The Partial Autocorrelation Function

The $AR(p)$ has an autocorrelation function which is infinite in extent, but it can be described by p non-zero functions of the autocorrelations. Denote by ϕ_{kj} , the j th coefficient in an autoregressive process of order k , so that ϕ_{kk} is the last coefficient we can obtain from

the Yule- Walker equations given below by equation (1)

$$\rho_j = \phi_1 \rho_{j-1} + \phi_2 \rho_{j-2} + \dots + \phi_p \rho_{j-p} \text{ for } j = 1, 2, \dots \quad (1)$$

which is generalised as

$$\rho_j = \phi_{k1} \rho_{j-1} + \phi_{k2} \rho_{j-2} + \dots + \phi_{k(k-1)} \rho_{j-k+1} + \phi_{kk} \rho_{j-k} \text{ for } j = 1, 2, \dots, k, \quad (2)$$

which may be written since $\rho_0 = 1$

$$\begin{pmatrix} 1 & \rho_1 & \rho_2 & \dots & \rho_{k-1} \\ \rho_1 & 1 & \rho_1 & \dots & \rho_{k-2} \\ \vdots & \vdots & \vdots & \dots & \vdots \\ \rho_{k-1} & \rho_{k-2} & \rho_{k-3} & \dots & 1 \end{pmatrix} \begin{pmatrix} \phi_{k1} \\ \phi_{k2} \\ \vdots \\ \phi_{kk} \end{pmatrix} = \begin{pmatrix} \rho_1 \\ \rho_2 \\ \vdots \\ \rho_k \end{pmatrix} \quad (3)$$

or in matrix notation as

$$\mathbf{P}_k \boldsymbol{\phi}_k = \boldsymbol{\rho}_k. \quad (4)$$

Solving these equations successfully for $k = 1, 2, 3 \dots$ we obtain

$$\begin{aligned} \phi_{11} &= \rho_1 \\ \phi_{22} &= \frac{\begin{vmatrix} 1 & \rho_1 \\ \rho_1 & 1 \end{vmatrix}}{\begin{vmatrix} 1 & \rho_1 \\ \rho_1 & 1 \end{vmatrix}} = \frac{\rho_2 - \rho_1^2}{1 - \rho_1^2} \\ \phi_{33} &= \frac{\begin{vmatrix} 1 & \rho_1 & \rho_2 \\ \rho_1 & 1 & \rho_1 \\ \rho_2 & \rho_1 & 1 \end{vmatrix}}{\begin{vmatrix} 1 & \rho_1 & \rho_2 \\ \rho_1 & 1 & \rho_1 \\ \rho_2 & \rho_1 & 1 \end{vmatrix}}. \end{aligned} \quad (5)$$

Alternatively

$$\begin{aligned}
 \rho_1 &= \phi_{21} + \phi_{22}\rho_1 \text{ or} \\
 \phi_{21} &= (1 - \phi_{22})\rho_1 \\
 \rho_2 &= \phi_{21}\rho_1 + \phi_{22} \\
 &= (1 - \phi_{22})\rho_1^2 + \phi_{22} \text{ or} \\
 \phi_{22} &= \frac{\rho_2 - \rho_1^2}{1 - \rho_1^2}.
 \end{aligned}$$

In general, for ϕ_{kk} , the determinant in the numerator has the same elements as that in the denominator, but with the last column replaced by ρ_k . The quantity ϕ_{kk} , regarded as a function of the lag k is called the partial autocorrelation coefficient. For an autoregressive process of order p , the partial autocorrelation function ϕ_{kk} will be nonzero for k less or equal to p and zero for k greater than p . In other words, the partial autocorrelation function of the p th order autoregressive process has a cut-off after lag p .

2.5.2 Portmanteau Test.

Financial applications often require to test jointly that several autocorrelations of r_t are zero. Box and Pierce(1970) proposed the Portmanteau statistic

$$Q^*(m) = T \sum_{l=1}^m \hat{\rho}_l^2 \quad (1)$$

as a test for

$$H_0 : \rho_1 = \dots = \rho_m = 0$$

vs

$$H_a : \rho_i \neq 0 \text{ for some } i \in \{1, \dots, m\}.$$

Under the assumption that $\{r_t\}$ is an *i.i.d.* sequence with certain moment conditions

$$Q^*(m) \approx \chi_m^2. \quad (2)$$

Ljung and Box(1978) modify the $Q^*(m)$ statistic as follows

$$Q(m) = T(T+2) \sum_{l=1}^m \hat{\rho}_l^2 / (T-l). \quad (3)$$

This increases the power in finite samples. Simulation studies suggest that the choice of $m \approx \log(T)$ provides better power performance. The function

$$\{\hat{\rho}_1, \hat{\rho}_2, \dots, \hat{\rho}_t, \dots\} \quad (4)$$

is called the *sample autocorrelation function* (*ACF*) of the log return r_t . It plays an important role in linear time series analysis. The sample *ACF* captures the dynamic time and stochastic structure of the time series process. (Troskie (2002))

2.6 White Noise and Linear Time Series.

If the series $\{r_t\}_{t=1}^T$ is a sequence of independent and identically distributed (i.i.d.) random variables with finite mean and variance

$$E(r_t) = 0, \text{ var}(r_t) = \sigma^2 \quad (1)$$

as a test for

$$H_0 : \rho_1 = \dots = \rho_m = 0$$

vs

$$H_a : \rho_i \neq 0 \text{ for some } i \in \{1, \dots, m\}.$$

Under the assumption that $\{r_t\}$ is an *i.i.d.* sequence with certain moment conditions

$$Q^*(m) \approx \chi_m^2. \quad (2)$$

Ljung and Box(1978) modify the $Q^*(m)$ statistic as follows

$$Q(m) = T(T+2) \sum_{l=1}^m \hat{\rho}_l^2 / (T-l). \quad (3)$$

This increases the power in finite samples. Simulation studies suggest that the choice of $m \approx \log(T)$ provides better power performance. The function

$$\{\hat{\rho}_1, \hat{\rho}_2, \dots, \hat{\rho}_t, \dots\} \quad (4)$$

is called the *sample autocorrelation function* (*ACF*) of the log return r_t . It plays an important role in linear time series analysis. The sample *ACF* captures the dynamic time and stochastic structure of the time series process. (Troskie (2002))

2.6 White Noise and Linear Time Series.

If the series $\{r_t\}_{t=1}^T$ is a sequence of independent and identically distributed (i.i.d.) random variables with finite mean and variance

$$E(r_t) = 0, \text{ var}(r_t) = \sigma^2 \quad (1)$$

then the series is called white noise. For a white noise series all the ACF' s are zero. If in addition,

$$r_t \sim N(0, \sigma^2) \quad (2)$$

then the series is called Gaussian white noise. A time series $\{r_t\}_{t=1}^T$ is said to be linear if it can be written as

$$r_t = \mu + \sum_{i=0}^{\infty} \psi_i e_{t-i}, \quad (3)$$

where $\psi_0 = 1$, $E(r_t) = \mu$ and $\{e_t\}_{t=1}^T$ is a white noise series with

$$E(e_t) = 0 \text{ and} \quad (4)$$

$$\text{var}(e_t) = E(e_t)^2 = \sigma_e^2.$$

The dynamic structure of $\{r_t\}_{t=1}^T$ is governed by the coefficients ψ_i which are called the ψ -weights. If $\{r_t\}_{t=1}^T$ is weakly stationary then

$$E(r_t) = \mu \text{ and}$$

$$\text{var}(r_t) = \sigma_e^2 \sum_{i=0}^{\infty} \psi_i^2.$$

The lag- l autocovariance of r_t is

$$\gamma_l = \text{cov}(r_t, r_{t-l}) \quad (5)$$

$$\begin{aligned} &= E \left[\left(\sum_{i=0}^{\infty} \psi_i e_{t-i} \right) \left(\sum_{j=0}^{\infty} \psi_j e_{t-l-j} \right) \right] \\ &= E \left[\sum_{i=0}^{\infty} \sum_{j=0}^{\infty} \psi_i \psi_j e_{t-i} e_{t-l-j} \right] \\ &= \sum_{j=0}^{\infty} \psi_{j+l} \psi_j E(e_{t-l-j}^2) \\ &= \sigma_e^2 \sum_{j=0}^{\infty} \psi_j \psi_{j+l}. \end{aligned}$$

Thus, the ψ -weights are related to the autocorrelations of r_t by

$$\begin{aligned}\rho_l &= \frac{\gamma_l}{\gamma_0}; \quad l \geq 0 \\ &= \frac{\sum_{i=0}^{\infty} \psi_i \psi_{i+l}}{1 + \sum_{i=1}^{\infty} \psi_i^2}, \quad \psi_0 = 1.\end{aligned}\tag{6}$$

In general, econometric and times series models are described by these ψ -weights. (Troskie (2002))

2.7 Validity of the Normal Assumption for forthcoming Analysis

Before performing any empirical investigation, a preliminary data analysis is undertaken to justify the basic premise on which maximum likelihood estimation and portfolio optimisation relies on. The fundamental assumption of normality for shares' log return is a crucial requirement for the portfolio mean-variance optimisation. Since the normal distribution is solely described by its mean and variance this is consistent with the mean-variance analysis and proves to be a convenient assumption. Fama (1976) stated that the assumption of normally distributed stock returns was an acceptable working approximation to returns on the NYSE, although Mandelbrot (1963) and Fama (1965) reported that stock returns were more leptokurtic than the normal distribution.

The data used in all the empirical analysis are monthly returns for all shares for the period July 1988 to February 2005, since they serve to compress the scale and hence converge to normality to a certain extent. Figure 2.4 below illustrates the monthly log return of the JSE stock Anglos.

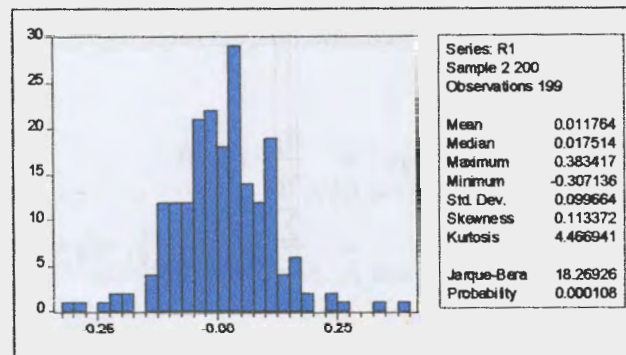
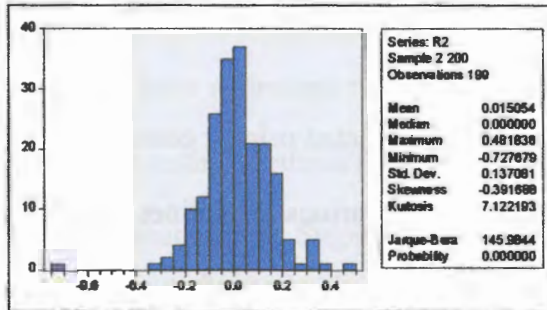


Figure 2.4: Monthly log Return of Anglo

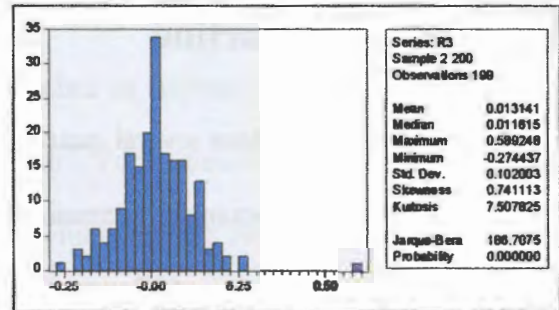
From a graphical stand point, at first glance it seems that the distribution is bell shaped and is quite close to normality and thus can be considered approximately normal. The skewness is positive and the kurtosis is greater than 3. The slight positive skewness, however is attributed to the positive returns over the period. The Jarque-Bera statistic compares and measures the difference in skewness and kurtosis of the series with that of a perfect normal distribution. The high significance of the statistic at the significance level suggests the rejection of the null hypothesis of normality. The assumption of normality is however reasonable for the purpose of this study. The reader is referred to appendix1 in the list of appendices section for histograms of the monthly log returns for the remaining 8 shares.

2.8 List Of Appendices

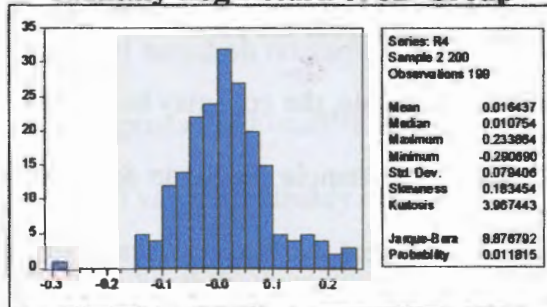
2.8.1 Appendix 1: Histogram of Monthly Log Returns of the remaining Eight Shares



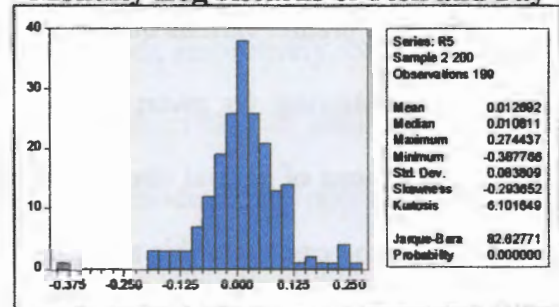
Monthly Log Return of JD Group



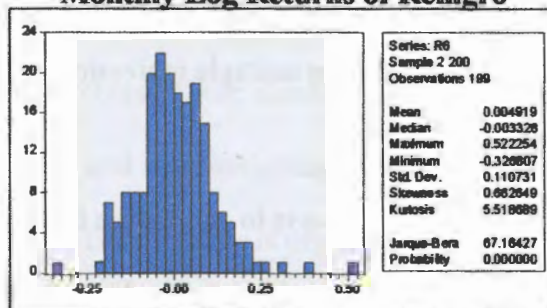
Monthly Log Returns of Pick and Pay



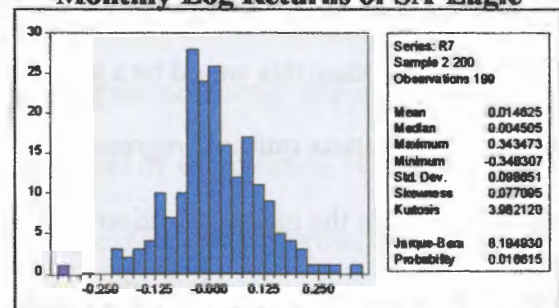
Monthly Log Returns of Remgro



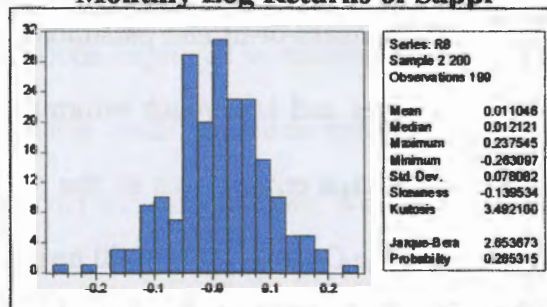
Monthly Log Returns of SA-Eagle



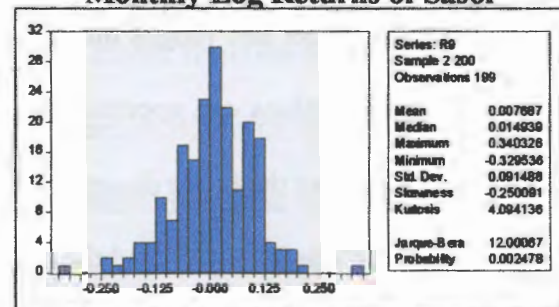
Monthly Log Returns of Sappi



Monthly Log Returns of Sasol



Monthly Log Returns of Tiger Brands



Monthly Log Returns of Tongaat

Chapter 3

Applied Regression Analysis and Model Selection

3.1 Introduction

Applications where several quantities are to be predicted using a common set of predictor variables are becoming increasingly important in various disciplines (Breiman & Friedman, 1997; Bilodeau & Brenner, 1999). For instance, in a manufacturing process one may want to predict various quality aspects of a product from the parameter setting used in the manufacturing. Or, given the mass spectra of a sample, the goal may be to predict the concentrations of several chemical constituents in the sample (Breiman & Friedman, 1997). Or as the case is in this study the object is to project the future prices of selected stocks on the JSE using two specific modelling approaches. A natural class of models that accommodate this would be a generalization of a univariate multiple regression model, called multivariate multiple regression (MMR).

In the model selection problem, the main objective is to choose the true model from a manageable set of candidate models. Information criterion gauges the validity of a statistical model and judges the balance between goodness-of-fit and parsimony; how well observed values can approximate to the true values and how much information can be explained by the lower dimensional model. Information criteria such as: the Akaike Information Criterion (AIC) and the Bayesian Information Criterion (BIC) will be introduced in

the following sections. Also, we present various improved criteria and intensive computing model selection criteria.

As mentioned above, the main objective is to choose the true model from a manageable set of competing models. Since statisticians devoted their attention to this problem, they have addressed two different approaches as follows: One way is to adapt the hypothesis testing methods to select the true model. For instance, Forward selection, Backward elimination, and Stepwise selection are included in this category. Forward selection and Backward elimination compute the F-statistic that reflects a variables contribution to the model based on no variables and all of the variables, respectively. Stepwise selection also computes the F-statistic for a model, but differs from the two methods above in the sense that variables already contained in the particular model do not necessarily stay there. However, the level of significance for hypothesis testing could be arbitrarily chosen by researchers and different models could be easily selected with the same data by different researchers. These disadvantages result from testing the sequential hypothesis in different order, and from assigning different choices of the level of significance. (Chatfield (1994))

This Chapter is organized as follows. In the next section we present the background to Applied regression analysis. In section 3, we present the earlier information criteria. Various improved versions of information criteria, model selection methods and other intensive model selection criteria are classified in section 4. In section 5, we present the Model Building process. We present an empirical study in the final section 6.

3.2 Applied Regression Analysis

When fitting a multiple linear regression model, a researcher will likely include independent variables that are not important in predicting the dependent variable Y . In the analysis he will try to eliminate these variable from the final equation. The objective in trying to find the best equation will be to find the simplest model that adequately fits the data. This will not necessarily be the model the explains the most variance in the dependent variable Y (the equation with the highest value of R^2). This equation will be the equation with all of the independent variables in the equation. Our objective will be to find the equation with the least number of variables that still explain a percentage of variance in the dependent variable that is comparable to the percentage explained with all the variables in the equation.

In multivariate multiple regression (MMR) q dependent variables (y_1, y_2, \dots, y_q) are to be predicted by linear relationships with k independent variables (x_1, x_2, \dots, x_k). In practice, MMR includes a large number of predictors where some of them might be slightly correlated with the y 's or they may be redundant because of high correlations with other x 's (Spark et al., 1985). The use of poor or redundant predictors can be harmful because the potential gain in accuracy attributable to their inclusion is outweighed by inaccuracies associated with estimating their proper contribution to the prediction (Spark et al., 1985).

The problem of determining the "best" subset of independent variables in multiple linear regression has long been of interest to applied statisticians, and it continues to receive considerable attention in recent statistical literature (McQuarrie & Tsai, 1998). Two approaches are suggested in the statistical literature to deal with this problem. The first ap-

proach is to find the "best" set of predictors for each individual response variable using one (or more) of the multiple model selection criteria that are available in most of statistical packages such as EVIEWS 3.1, Ox etc.

Sparks et al. (1985) criticized univariate model selection methodology as compared to multivariate techniques and stated two reasons for dealing with target variables jointly rather than separately. One reason is simply that it is computationally more efficient because the number of times required doing necessary computations for model selection would be reduced from q to one. A second reason is that researchers sometimes need to establish which subset of predictors can be expected to perform well for all target variables, especially if there are costs associated with sampling the predictors.

3.2.1 Assumptions underlying the Multiple Linear Regression Model.

The set of assumptions placed on the model and data in order for the multiple regression model to work is similar to that for the simple regression model. The only modifications involve taking into account the existence of more than one independent variable in the relationship

1. Linearity of the Model: The regression model is linear in the unknown coefficients, i.e.

$$Y_t = \beta_0 + \beta_1 X_{1t} + \beta_2 X_{2t} + \dots + \beta_k X_{kt} + e_t = \beta_0 + \sum_{i=1}^k \beta_i X_{it} + e_t \text{ for all } t = 1, 2, \dots, N.$$

2. Errors average to Zero: The error term e_t is a random variable with mean zero, i.e. $E[e_t] = 0$.

3. Variability in the X's: Not all the observations on each X_i are identical for all $i = 1, 2, \dots, k$, at least one is different

for each X_i .

4. Non-stochastic Regressors: X_t is given and non-random, it is therefore uncorrelated with the error term, i.e.

$$\text{cov}(X_{it}, e_t) = 0 \text{ for all } i = 1, 2, \dots, k.$$

5. Homoskedasticity: e_t has constant variance for all t , i.e. $\text{Var}(e_t) = E(e_t^2) = \sigma^2$

6. Serial Independence: e_t and e_s are independently distributed for all $t \neq s$, so that $\text{Cov}(e_t, e_s) = 0$.

7. Normality of Errors: e_t is normally distributed so that $e_t \sim N(0, \sigma^2)$, which implies that

$$Y_t \sim N\left(\sum_{i=1}^k \beta_i X_{it}, \sigma^2\right) \text{ for given independent variables } \{X_{it}\}_{i=1}^k.$$

3.2.2 Method of Moments and Least Squares

Suppose that the assumptions 1-7 made above hold and that we observe a sample $\{y_i, x_i\}$ for $i = 1, \dots, n$. Consider the problem of estimating the parameters β and σ^2 of the linear model:

$$y_i = x_i' \beta + \varepsilon_i, \quad i = 1, \dots, n$$

$$E[\varepsilon_i^2 | X] = \sigma^2$$

The method of least squares determines the estimates of β by minimizing the sum of squares residuals (SSR) given by:

$$\begin{aligned}
 SSR(\tilde{\beta}) &= \sum_{i=1}^n \tilde{\varepsilon}_i^2 \\
 &= \sum_{i=1}^n (y_i - x_i' \tilde{\beta})^2 \\
 &= (y - X\tilde{\beta})'(y - X\tilde{\beta}) \\
 &= \tilde{\varepsilon}'\tilde{\varepsilon}
 \end{aligned}$$

where $\tilde{\beta}$ is an arbitrary value of β . Let $\hat{\beta}$ denote the least squares estimate of β . Then $\hat{\beta}$ is defined as:

$$\hat{\beta} = \arg \min_{\tilde{\beta}} SSR(\tilde{\beta})$$

to determine b , note that

$$SSR(\tilde{\beta}) = y'y - 2y'X\tilde{\beta} + \tilde{\beta}'X'X\tilde{\beta}$$

then

$$\frac{\partial SSR(\tilde{\beta})}{\partial \tilde{\beta}} = -2X'y + 2X'X\tilde{\beta} \text{ (normal equations)}$$

setting this derivative equal to zero (solving the normal equations) and solving for $\hat{\beta}$ gives:

$$\hat{\beta} = (X'X)^{-1}X'y$$

Remark

$\hat{\beta}$ may be expressed in terms of sample moments of the data

$$\begin{aligned}\hat{\beta} &= \left(\frac{1}{n} X'X \right)^{-1} \frac{1}{n} X'y \\ &= \left(\frac{1}{n} \sum_{i=1}^n x_i x_i' \right)^{-1} \frac{1}{n} \sum_{i=1}^n x_i y_i \\ &= S_{xx}^{-1} S_{xy}\end{aligned}$$

The least squares fitted values are

$$\hat{y} = X\hat{\beta}$$

or

$$\hat{y}_i = x_i' \hat{\beta}$$

The least squares residuals are defined as

$$\begin{aligned}e &= y - X\hat{\beta} \\ &= y - \hat{y}\end{aligned}$$

or

$$\begin{aligned}e_i &= y_i - x_i' \hat{\beta} \\ &= y_i - \hat{y}_i\end{aligned}$$

Notice that the least squares residuals solve the normal equations

$$\begin{aligned}X'(y - X\hat{\beta}) &= X'e = 0 \\ \implies X &\text{ is orthogonal to } e\end{aligned}$$

Method of Moments Interpretation of Least Squares

There is a simple method of moments interpretation of least squares. Assumption 2 gives the population moment restriction

$$E[x_i \varepsilon_i] = 0, \quad i = 1, \dots, n$$

which may be expressed as

$$E[X' \varepsilon] = 0$$

or

$$E[X'(y - X\beta)] = 0$$

The method of moments estimator determines the estimate of β by making the appropriate sample moment satisfy the population moment restriction:

$$\begin{aligned} \frac{1}{n} X'(y - X\hat{\beta}) &= S_{xy} - S_{xx}b = 0 \\ \implies \hat{\beta} &= (X'X)^{-1}X'y \end{aligned}$$

3.2.3 Properties of the OLS Estimators in the Multiple Regression Model

Properties of estimators in the multiple regression model are similar to those derived for simple regression. It is possible to prove that the estimates $\{\hat{\beta}\}_{i=1}^k$ have the following properties:

- OLS Estimators $\{\hat{\beta}\}_{i=1}^k$ are Unbiased.
- OLS Estimators $\{\hat{\beta}\}_{i=1}^k$ are Consistent so long as $Cov(X_{it}, \varepsilon_i) = E(X_{it}, \varepsilon_i) = 0$,

and in addition

$$0 < Var(X_i) < \infty \quad (X_i \text{ is stationary}).$$

• OLS Estimators $\{\hat{\beta}\}_{i=1}^k$ are Best Linear Unbiased Estimators (BLUE), that is, the OLS estimators have the

lowest variance among any other estimators that are linear functions of the data.

one can show that an unbiased estimator of the variance of the stochastic error term is:

$$s^2 = \hat{\sigma}^2 = \frac{\sum \hat{\varepsilon}_i^2}{n - k}$$

which is the sum of squared fitted errors divided by $n - k$ because k degrees of freedom are lost estimating $\{\hat{\beta}\}_{i=1}^k$

3.2.4 Properties of the OLS Estimators in the Simple Regression Model

Assumption 7. states that the stochastic error term is distributed normally, and it can be shown that the distributions for the OLS estimators are:

$$\begin{aligned} \hat{\beta}_i &\sim N(\beta_i, \sigma_{\hat{\beta}_i}^2) \text{ where } \sigma_{\hat{\beta}_i}^2 = \text{Var}(\hat{\beta}_i) = E(\hat{\beta}_i - \beta_i)^2 \text{ for all } i = 1, 2, \dots, k \\ \frac{\sum \hat{\varepsilon}_i^2}{\sigma^2} &= \frac{(n - k)\hat{\sigma}^2}{\sigma^2} \sim \chi_{N-k}^2 \end{aligned}$$

from the standpoint of how precisely $\{\hat{\beta}\}_{i=1}^k$ estimate the true parameters, it can again be shown that the larger is the sample size and the larger the variation in the X's, the smaller will be the variance of the coefficient estimates. Hence the OLS estimates are more precise given a larger number of observations and the higher the variability in the explanatory variable X. It should be made aware of the fact that it is not possible to estimate the model if the number of observations is less than the number of coefficients. (Chatfield (1994))

3.2.5 All Subsets Regression

Consider the linear model

$$Y = \beta_0 + \beta_1 X_1 + \cdots + \beta_p X_p + e. \quad (1)$$

In variable selection procedures we simply want to select important or relevant variables, and discard unimportant or irrelevant variables. One of the best procedures is the all subsets procedure. When mainframe computers were very expensive this method was very seldom used since it was prohibitively expensive, especially for models with a large set of $X_variables$. With computer time freely available on PC's and inexpensive, even for very large models, the method has become very popular and is the recommended procedure.

As the procedure suggests we simply fit all subsets, which includes the full set. We need to find a criterion on which to base our selection. Two measures immediately suggest themselves, namely the multiple correlation coefficient R^2 , and the adjusted multiple correlation coefficient R_{adj}^2 . Thus we have

$$R^2 = \frac{\sum_{i=1}^n (\hat{Y}_i - \bar{Y})^2}{\sum_{i=1}^n (Y_i - \bar{Y})^2} = \frac{\text{Explained variation}}{\text{Total variation}}$$

where $100R^2$ explains the amount of variation in the model. Thus we select models with large R^2 . The adjusted R_{adj}^2 is

$$R_{adj}^2 = 1 - \frac{SSE/(n-k)}{SST/(n-1)} = 1 - \frac{\frac{SSE}{(n-k)}}{\frac{SST}{(n-1)}},$$

and is similar to R^2 except that it adjusts for the degrees of freedom. Thus it is ideally suited to compare different models. This is the measure recommended for the all subsets selection procedure. The algorithm is as follows.

Algorithm for All Subsets Regression.

a $F_statistic$ which is well-known to us. Consider again our model

$$Y = \beta_0 + \beta_1 X_1 + \cdots + \beta_p X_p + e.$$

For any of the estimates of β_i we can compute the statistic $t_i = \frac{\hat{\beta}_i}{\sqrt{s^2 c_{ii}}}$. In variable selection procedures we rather use the $F_i_statistic$ which is given by

$$F_i = t_i^2 = \frac{\hat{\beta}_i^2}{s^2 c_{ii}} \sim F_{1, n-k}. \quad (1)$$

As mentioned before these F_i or t_i statistics are dependent, and it is difficult to control the Type 1 error if several such tests are performed. In stepwise procedures we do not perform tests, but rather use the $F_i_statistics$ as diagnostics or *selection* criteria for building the models.

We usually choose two values for the $F_statistics$ called F_to_enter or F_{in} and F_to_delete or F_{out} . For F_{in} we usually choose a value between 1 and 4, since the value 4 roughly corresponds to a $|t|$ -value of 2 (that is $t^2 = 4$). The larger the F_value the fewer variables are selected. Choosing a very small F_value virtually brings in all variables in the equation. For the F_{out} we usually choose a value slightly less than the F_{in} . The reason being that the F_{out} discards variables after they have been selected, and one is reluctant to do this, if a variable has been selected at a previous step. We will discuss this rational in more detail below.

We now discuss separately the forward and backward selection procedures.

3.3.2 Forward Selection Procedure.

The forward selection procedure proceeds in the following steps.

Step 0.

In this first step the F_i statistics are computed for all the variables in the equation. This step is the same as the first step of the all subsets procedure. The largest of these F_i statistics is selected. If this largest F_i is larger than F_{in} then it is the first variable entered in the equation. If the largest F_i value is less than F_{in} then even the most important variable is not selected and the procedure stops. That is, we have no model and the best estimate for Y is \bar{Y} .

Assume that the largest F_i is larger than F_{in} , and for convenience sake say it is F_1 so that variable X_1 is selected in the model. This largest F_i is also the largest simple correlation with Y . In other words since we have selected X_1 the correlation between Y and X_1 given by r_{yx_1} is the largest. The relationship between the F statistic and the correlation coefficient r_{yx_1} is

$$F = (n - 2) \frac{r_{yx_1}^2}{1 - r_{yx_1}^2}. \quad (2)$$

Clearly the largest F is also the largest correlation r .

Step 1.

In this step the first variable, - in our case - X_1 , is brought in the model and the F_i statistics for all the variables not in the model are computed. The largest F_i is chosen. Say, for convenience sake, it is X_2 . If this largest F_i statistic (now it is F_2) is larger than F_{in} , then it is selected in the model. If not, then the process is stopped, and only X_1 is selected for the model.

This largest F_i statistic is also the one with the largest partial correlation. Since we have assumed that it is X_2 , what we are saying is that the partial correlation between Y

and X_2 , given that X_1 is in the equation, is the largest. What is the partial correlation? If we have the conditional distribution of the pair of random variables (Y, X_2) given $X_1 = x_1$ then the conditional correlation between (Y, X_2) given x_1 is called the partial correlation. If this partial correlation is called $r_{yx_2.x_1}$ then the relation with F is

$$F = (n - 3) \frac{r_{yx_2.x_1}^2}{1 - r_{yx_2.x_1}^2}.$$

Clearly the largest F is also the largest partial correlation coefficient. At this point we would just like to say that the partial correlation is very popular amongst Psychologists and Psychometricians and often used for selection procedures. It is therefore useful to know the relationship with the F statistic, which is normally used for selection procedures.

Step 2.

The process is repeated. The variables X_1 and X_2 is in the model. Choose as next variable the one with the largest F_i . If this largest F_i is larger than F_{in} then select the variable and continue. Otherwise stop and only the two variables (X_1, X_2) is in the model. Thus we select the variables of the previous step.

Step 3.

This step is very important in the forward selection procedure. At first it proceeds exactly as the previous steps and selects the next important variable provided that the F statistic (being the largest) is larger than F_{in} . Suppose it is X_3 . We now have in the equation (X_1, X_2, X_3) and the estimated model is

$$\hat{Y} = \hat{\beta}_0 + \hat{\beta}_1 X_1 + \hat{\beta}_2 X_2 + \hat{\beta}_3 X_3 \quad (3)$$

The forward procedure now looks back. It compares the F_i statistics with F_{out} . If it is found that one of the F_i -which will now be the smallest- is less than F_{out} then this variable is now deleted. Say, for convenience sake, it is F_2 , that is $F_2 < F_{out}$, then the variable X_2 is deleted and the model is

$$\hat{Y} = \hat{\beta}_0 + \hat{\beta}_1 X_1 + \hat{\beta}_3 X_3$$

We now continue with further stepping. Why do we look back. Simply our variables (X_1, X_2, \dots, X_p) themselves are correlated(not orthogonal) and variables which appear to be important in the beginning stages may become redundant as more variables are added. Partial correlations play an important part here.

Step 4.

All further steps now proceed as in Step 3. **We look forward and backward.** Variables are added if the F_i statistics are larger than F_{in} and they are deleted if they are less than F_{out} .

The model finally selected is when at a new step none of the remaining variables meets the F - to - enter (F_{in}) criterion and the model of the previous step is chosen.

Once variables have been selected at some stage in the model, then one is reluctant to discard them. For this reason the F_{out} is chosen slightly less than F_{in} . For example if $F_{in} = 2$ then select $F_{out} = 1.5$.

Comment.

The forward selection procedure is a very good procedure. Choosing a large F to enter a decision could be reached very soon. On the other hand if the experimenter wants to see as many variables in the equation, together with the importance in which they enter,

then a small value of F -to-enter can be chosen. The forward selection's final step can also be compared with the all subsets approach. The final forward selection model must be one of the sub models of the all subsets models. Comparing the different adjusted R_{adj}^2 one may well choose a model that is satisfactory from an economic point of view, but not optimal.

3.3.3 Backward Selection Procedure.

The backward selection starts with all variables in the equation. Only an F_{out} is chosen or $F_{in} = F_{out}$.

Step 0.

All variables are in the equation. The F statistics (and partial correlations) are computed. Each variable is considered as if it is the last variable entering the equation. The smallest F_i statistic is compared with the F_{out} . If the smallest F_i is larger than F_{out} then all variables are important and the method stops. If the smallest F_i is less than F_{out} then this variable is discarded and we go to step 1.

Step 1.

The process is the same. All variables except the one discarded is in the model. The method either stops if the new smallest F_i is larger than F_{out} or continues with the next step if the smallest F_i is less than F_{out} in which case this last variable is discarded.

Stepping thus continues until at the final step no more F_i statistics are less than F_{out} .

Comment.

This is a bad procedure and not to be recommended. A problem that we have not tackled in this course is the problem of collinearity. Simply stated we assume there is

a linear relationship between Y and the X 's. There may well be, and more often than not, there are linear relationships between the X variables. Such linear relationships are called collinearities. Such collinearities can have disastrous effects on the OLS and MLE estimates. A sure indication of collinearities are large confidence intervals for some of the β_i compared to others.

It is well known, that because of collinearities, that the backward procedure can give entirely different results from the forward selection procedure. There are remedies but they are beyond the scope of this course. However, there are two obvious solutions. The all subsets procedure, where we start with a few variables and build up the model. The chances are very good that we will select a model early with few X variables in the equation to avoid the bad effect of collinearity. The forward selection procedure is also very good to combat the effect of collinearity, especially if the F - to enter is chosen large, say between 3 and 4.

3.3.4 The C_p criterion of Mallows

Another criterion which is often used for variable selection is the C_p criterion of Mallows(a past UCT statistics student). The C_p criterion is especially popular amongst the Ivy League Universities in the U.S.A. The C_p statistics is

$$C_p = \frac{SSE(p)}{s^2} - (n - 2p) \quad (4)$$

where $SSE(p)$ is the error or residual sums of squares containing p parameters β_i including the constant term. The estimate s^2 is the estimate of σ^2 from the largest equation postulated containing all the variables, which is k in our case. The C_p statistic is closely related to R_{adj}^2 and may be preferred to R_{adj}^2 . Now if an equation with p parameters is

adequate, that is, it represents a good fit then

$$E(SSE(p)) \approx (n - p)\sigma^2 \quad (5)$$

Because we are also assuming that $E(s^2) = \sigma^2$, it is true *approximately*, that

$$\begin{aligned} E(C_p) &= \frac{E(SSE(p))}{\sigma^2} - (n - 2p) \\ &= \frac{(n - p)\sigma^2}{\sigma^2} - (n - 2p) = p \end{aligned} \quad (6)$$

for an adequate model. It follows that the plot of C_p versus p will show up the 'adequate models' as points fairly close to the $C_p = p$ line. In general we would look for models with low p where C_p is close to p .

3.3.5 Information Criteria and Complexity.

There are several information criteria available to determine the order p of a model(p) of a regression model or times series process. All of them are likelihood based. For example the well-known Akaike Information Criterion (Akaike (1973)) is defined as

$$AIC = -\frac{2l}{T} + \frac{2}{T} \times (\text{number of parameters})$$

where the likelihood is evaluated at the maximum likelihood estimates (m.l.e.) and T is the sample size, that is

$$\begin{aligned} l &= \log(\text{likelihood}) \\ &= -\frac{T}{2}(1 + \log 2\pi + \log(\hat{\epsilon}'\hat{\epsilon}/T)). \end{aligned}$$

For a Gaussian $AR(k)$ model, the AIC reduces to

$$AIC(k) = \log \hat{\sigma}_l^2 + \frac{2k}{T}$$

where $\hat{\sigma}_T^2$ is the maximum likelihood estimate of σ^2 which is the variance of e_t . In practice one computes the $AIC(k)$ for $k = 1, \dots, p$ and selects the one that has the minimum AIC value. The term $2k/T$ is called the penalty function because it penalizes a candidate model by the number of parameters used. Different penalty functions result in different information criteria. The Schwarz Criterion (SC) is an alternative to the AIC that imposes a larger penalty for additional coefficients

$$SC = \frac{-2l}{T} + (k \log T)/T$$

The Schwarz Criterion has been shown to select the best models in a Regression and Times Series environment by A.E Clark (2003).

3.3.6 The Model Building Process

There are a number of steps in building an adequate regression model.

1. Collect the data.

(a) Controlled Experiment or Observed Study?

The collection of data depends very much on whether you (the statistician) planned it out (controlled experiment) or whether this data was given to you and you are trying to make the best of it (observed study). Typically, more information can be derived from a controlled experiment, rather than an observed study, because you can set key quantities, such as sample size, or design the experiment in such a way as to anticipate problems related to multi-collinearity or nonlinearity or non-normality, for example.

(b) Data diagnostics.

However the data is collected, diagnostic procedures such as scatter plots and correlation matrices should be applied to the data to find out not only if there are problems such as outliers, non{constant variance or multicollinearity in the data, but also to provide direction, if necessary (in observed studies), as to what model (what explanatory variables) to use to fit the data.

(c) Data remedial measures.

Remedial measures such as transforming the data or (carefully) eliminating outliers should be used to try and fix any problems identified in the *previous data diagnostic stage*.

Another round of data diagnostics should then be applied to the ~~revised data set~~.

2. Develop a model to fit the data.

(a) Reduction of explanatory variables.

In observed studies, it is often necessary to reduce the number of explanatory variables to a more manageable (three or four, say) number. After this is done, there is often several models with different explanatory variables to choose from.

(b) Model/Data diagnostics.

Diagnostic procedures such as residual plots, scatter plots and correlation matrices should be used to determine not only how well the model(s) fit(s) the data but also, if necessary (in observed studies), to choose a "best" model.

(c) Model/Data remedial measures.

Remedial measures such as introducing curvature or interaction effects, or transforming the data should be used to try and fix any problems identified in the previous model/data diagnostic stage. Another round of model/data diagnostics should then be applied to the

revised model.

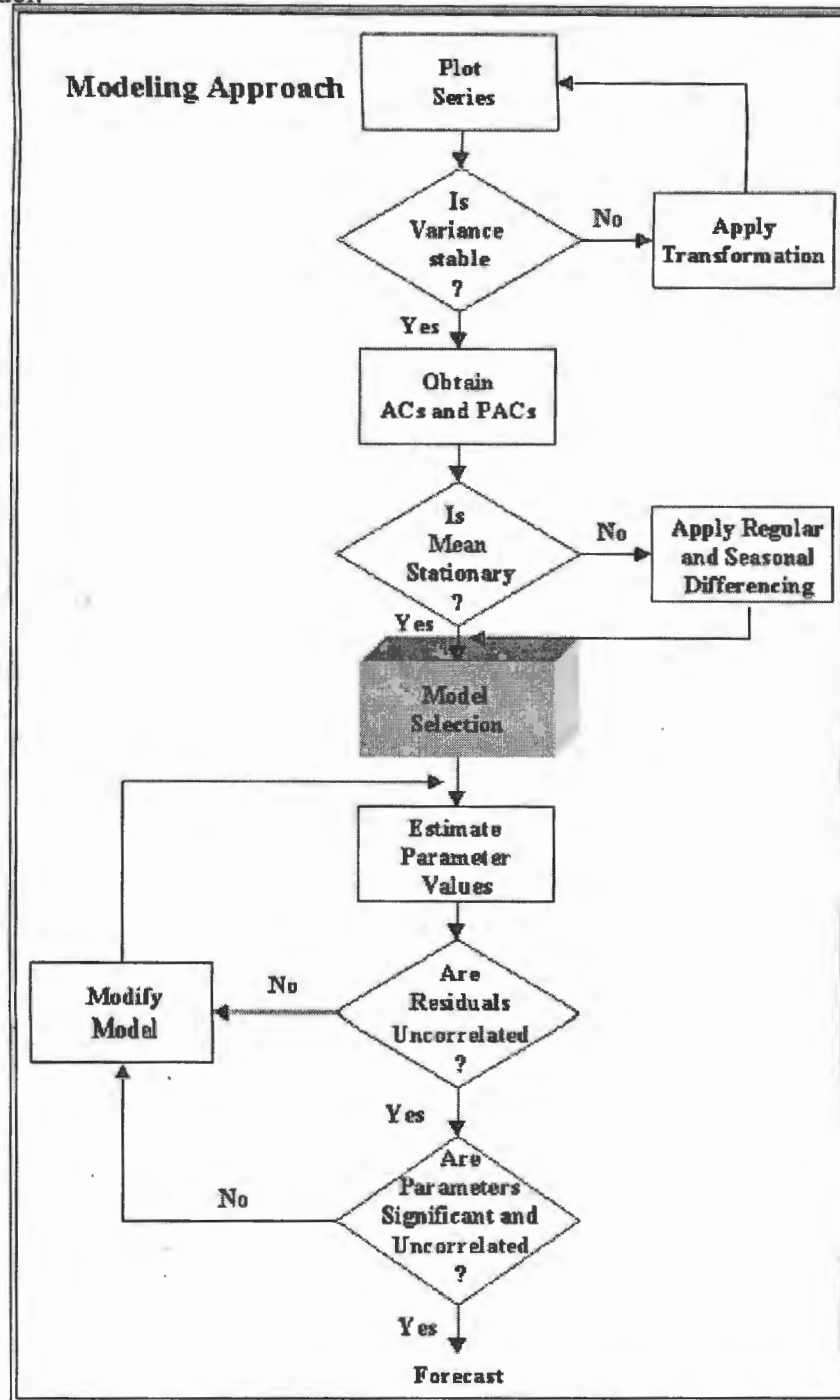


Figure 3.1: The Model Building Process

How do we find a suitable model? As illustrated above, by figure 3.1, the answer depends

on various considerations, including the properties of the series as assessed by a visual examination of the data, the number of observations available, the context and the way the model is to be used.

It is important to understand that model building has three main stages, which can be described as:

- (1) Model Formulation (or model specification)
- (2) Model estimation (or model fitting)
- (3) Model checking (or model verification).

Most literature often concentrate on estimation, but barely focus about the more important topic of forecasting the model. This is unfortunate because modern time series software makes model fitting straightforward for numerous types of models, so the real problem is knowing which model to fit in the first place.

Model checking is also of vital importance, and the assessment of residuals is an essential step in the analysis. Once again modern software makes this relatively painless and may result in an initial model being discredited. Then alternative models are tested. Usually there are several cycles of model fitting as a model is modified and improved in response to residual checks or in response to additional data. Thus model building is an iterative and interactive process.

With regards to model formulation, the analyst should consult appropriate 'experts' about the given problem, ask questions to get relevant background knowledge, look at a time plot of the data to assess their more important features, and make sure that a proposed model is consistent with empirical, theoretical and with the objectives of the investigation.

In many areas, especially with economics and finance, non-stationary series often arise and in addition may be fairly short. This problem is quickly overcome by differencing the observed time series until it becomes stationary and then fitting an appropriate model to the differenced series.

Sometimes the analyst may have several competing models in mind then it may help to look at a model-selection statistic such as the Akaike's Information Criterion (*AIC*). These statistics try to strike a balance between the need for a 'parsimonious' model, which uses as few parameters as possible, and a model that is too simple and overlooks important effects. A useful reference on model building, in general, and model selection statistics and the Principle of Parsimony, in particular, is Burnham and Anderson (2002).

Whatever model is fitted, it is important to realise that it is only an approximation to the 'truth', and the analyst should always be prepared to modify a model in light of new evidence. When searching for a model, it is common to try many different models. Although statistics, like the *AIC* and *BIC*, penalize more complex models, it should be realised that there is still a danger that fitting many models to the same data may give a spuriously complex model that appears to give a good fit, but which nevertheless gives poor out-of-sample predictions/forecasts.

When a model is selected using the data, the analyst needs to remember that:

- (1) the true model may not have been selected,
- (2) the model may be changing through time, or
- (3) there may not be a true model anyway.

It is indeed strange that we often implicitly admit that there is uncertainty about the underlying model by searching for a 'best fit' model, but then ignore this uncertainty when making predictions. In fact it can readily be shown that, when the same data are used to formulate and fit a model, as is typically the case in time series analysis, then least squares theory does not apply. Parameter estimates will typically be biased. In other words, the properties of an estimator may depend not only on the selected model but also on the selection process.

It has been found in practice and by this study, as we shall see later, that out-of-sample forecast accuracy is generally (much) worse than would be expected from the within-sample fit of time-series models. As a result, prediction intervals tend to be too narrow in that 95% prediction intervals will typically contain fewer than 95% of future observations. Many analysts think that a narrow interval is somehow 'good', but theory suggests that it is safer to construct a wider interval that properly reflects model uncertainty as well as other sources of variation.

How then may we begin to cope with model uncertainty? Perhaps the most important step is to realise that the fitted model should be thought of as a useful approximation, and that our assessment of uncertainty is likely to be an underestimate. If we restrict ourselves to using a single best-fit model, then bear in mind that:

(1) a local model which changes through time may be preferred to a global model with constant parameters;

(2) a simple model may be preferred to a complicated model, even if the latter appears to fit better;

(3) a robust model may be preferred to a model that is optimal for one set of conditions only.

However, instead of identifying a single model to utilise, it is often worth considering the use of more than one model, especially when several different models appear to fit a set of data about equally well. When it comes to the issue of forecasting, an approach is to combine forecasts from several different methods and models by taking some sort of weighted average. An alternative to this is conducting a scenario analysis which produces a range of forecasts based on different, clearly stated model assumptions. (Mills (1999))

3.4 Empirical Study: Single Index

3.4.1 Introduction

When fitting a multiple linear regression model, a researcher will most likely include independent variables that are not important in predicting the dependent variable Y . In the analysis one will try to eliminate these variables from the final equation. The objective in trying to find the "best" equation will be to find the simplest model that adequately fits the data. This will not necessarily be the model that explains the most variance in the dependent variable Y (the equation with the highest value of R^2). Obviously this model will be the equation with all of the independent variables in the model. Our objective will be to find the equation with the least number of variables that still explain a percentage of variance in the dependent variable that is comparable to the percentage explained with all the variables in the equation.

In the mean-variance frontier build up for the Sharpe Single Index model, this section presents the regression models for our financial portfolio of nine stocks for the single index scenario. The index used for this case will be the JSE All Share Index which will be referred to as JSE in the remainder of this section.

3.4.2 The data

All of the empirical studies relate to financial stock market data. For this empirical analysis, the data used were the monthly log returns for the stocks Anglos, Jdgroup, Pick and Pay, Remgro, SA-Eagle, Sappi, Sasol, Tigerbrands and Tongaat. The modelling was done over the period July 1988-February 2005. Since the focus of this study is on the single index model, we have only one explanatory variable for each of our nine stocks, namely JSE. The data for JSE also spans over the same period.

The data was graciously obtained from Professor Cas Troskie (Supervisor) at the University of Cape Town, Department of Statistical Sciences. The returns used were non-overlapping returns due to two reasons:

- Investors usually deal with non-overlapping returns
- And secondly to avoid autocorrelation in the data.

It was decided to concentrate on real returns, exclusive of dividends. It seemed more realistic and logical to focus on these type of returns, considering that in the financial world dividend payouts are usually irregular and rare. Also most financial professionals tend to deal with real returns rather than nominal. It should be noted to the reader that the data pertaining to this study will also be applicable to the forthcoming single index empirical

studies performed in the remaining other chapters. Continuous reference will be made to this data as the Single Index Data Set (SIDS)

3.4.3 Study Objectives

The purpose of the study in this section is to build single index regression models for the stocks Anglos, Jdgroup, Pick and Pay, Remgro, SA-Eagle, Sappi, Sasol, Tigerbrands and Tongaat (represented as $R_1, R_2, R_3, R_4, R_5, R_6, R_7, R_8$ and R_9 respectively). using JSE as the explanatory variable defined in the 'The data' sub-section. Our overall objective will be to build a mean-variance frontier using parameters from the single index regression equation, which will be established in later chapters.

There is no significant reason for choosing these nine specific stocks for our portfolio, only with the overall objective in mind of showing the predominant frontier of the methods to follow. In the subsequent time series, state space and Garch chapters, we develop similar models for our portfolio in order construct efficient frontiers of these models. Thus the objective of this empirical study is to construct simple single index regression models for all nine stocks in our portfolio.

3.4.4 Methodology

The response variables Anglos, Jdgroup, Pick and Pay, Remgro, SA-Eagle, Sappi, Sasol, Tigerbrands and Tongaat., shall be represented as $R_1, R_2, R_3, R_4, R_5, R_6, R_7, R_8$ and R_9 respectively. The explanatory variable is the JSE All Share Index which we shall represent

as JSE. The programming language utilised to employ the computations and model building processes is the Time Series Package "EViews 5".

The model building method adopted in this study for all four responses is the Forward Selection process. For each of the nine stocks, JSE will be fitted as an explanatory variable. P -values in the range of 0.05 to approximately 0.12 will be regarded as an acceptable range of significance (i.e. 5% – 12%).

The principal criteria involved in deciding on these models are the R^2 , $Adj-R^2$, AIC , BIC , the F -statistic and it's corresponding Probability(P -value). The following sub-section discloses the results realised from adopting the Forward Selection process and the application of the principal selection criteria.

3.4.5 Primary Findings

Implementing the methodology outlined in the preceding section, table 3.1 below summarises the results from fitting the JSE as a market proxy in the nine single index regression models.

Regression Models	Model Selection Criteria						
	Adj- R^2	S.E. of reg	D-W stat	AIC	BIC	F-stat	Prob(F-stat)
Anglos = $Const + \beta_1 JSE + e_1$	0.6853	0.0559	2.0439	-2.9201	-2.8870	432.1429	0.0000
Jdgroup = $Const + \beta_2 JSE + e_2$	0.1955	0.1230	2.1197	-1.3440	-1.3109	49.1156	0.0000
Pick and Pay = $Const + \beta_3 JSE + e_3$	0.1514	0.0940	2.5776	-1.8817	-1.8487	36.3130	0.0000
Remgro = $Const + \beta_4 JSE + e_4$	0.4043	0.0613	2.2773	-2.7365	-2.7034	135.3852	0.0000
SA-Eagle = $Const + \beta_5 JSE + e_5$	0.0434	0.0820	1.9712	-2.1549	-2.1218	9.9893	0.0018
Sappi = $Const + \beta_6 JSE + e_6$	0.3276	0.0908	1.8826	-1.9503	-1.9172	97.4576	0.0000
Sasol = $Const + \beta_7 JSE + e_7$	0.3661	0.0785	1.8729	-2.2403	-2.2072	115.3412	0.0000
Tigerbrands = $Const + \beta_8 JSE + e_8$	0.2640	0.0670	2.1437	-2.5587	-2.5256	72.0310	0.0000
Tongaat = $Const + \beta_9 JSE + e_9$	0.2129	0.0812	2.0225	-2.1747	-2.1416	54.5664	0.0000

Table 3.1: Summary of Results for the 9 single index regression models

Investigating the results in table 3.1 above, it is immediately clear that all nine regression models fitted are exceptionally significant at the 5% level of significance. Secondly it is

also observed that Anglos has the highest R^2 , 68.53% followed by Remgro with 40.43% right up to SA-Eagle which exhibits the lowest of the nine. This can be attributed to the fact that the JSE all share index is comprised of all nine of the above shares and more. Since this index is value weighted and is constructed in terms of market capitalisation, it is obvious that Anglos would have the largest R^2 , as it has the highest market capitalisation on the JSE. Thus of the nine, SA-Eagle has the smallest market capitalisation.

The R^2 statistic is an indication of the amount of variation in the dependant explained by the model's explanatory variables. From above, it can be concluded that the JSE explains most of the variation in all nine shares except for SA-Eagle, where it only explains a mere 4%.

Furthermore, the Durbin-Watson (D-W) statistic is especially greater than 2 for R_1 , R_2 , R_3 , R_4 , R_8 and R_9 . This means that there might be a strong presence of negative autocorrelation present. As a consequence time series components will be required to be investigated in the subsequent chapter for these stocks. The converse is true for the remaining three.

Nearly all nine stocks have relatively stable standard error (SE) of regression except for R_2 , which is the only stock that exhibits quite a large SE when compared with the rest. Table 3.2 below provides a summary of the beta coefficient statistics of JSE for all nine models fitted.

Regression Models	Coefficient of JSE			
	Coefficient (β_i)	Std. Error	t-Statistic	Prob.
Anglos=const+ β_1 'JSE+ e_1	1.3973	0.0672	20.7880	0.0000
Jdgroup=const+ β_2 'JSE+ e_2	1.0359	0.1478	7.0083	0.0000
Pick and Pay=const+ β_3 'JSE+ e_3	0.6807	0.1130	6.0260	0.0000
Remgro=const+ β_4 'JSE+ e_4	0.8573	0.0737	11.6355	0.0000
SA-Eagle=const+ β_5 'JSE+ e_5	0.3115	0.0985	3.1606	0.0018
Sappi=const+ β_6 'JSE+ e_6	1.0776	0.1092	9.8721	0.0000
Sasol=const+ β_7 'JSE+ e_7	1.0141	0.0944	10.7397	0.0000
Tigerbrands=const+ β_8 'JSE+ e_8	0.6835	0.0805	8.4871	0.0000
Tongaat=const+ β_9 'JSE+ e_9	0.7208	0.0976	7.3869	0.0000

Table 3.2: Coefficient Statistics

Observing Table 3.2 above, the t-statistics and p-values for each of the regressions β estimates are highly significant. The stocks that hold the highest t-stats are the ones with the considerably lower standard errors. Namely Anglos, Remgro and Sasol have exceptionally high t-statistics.

3.4.6 Conclusions

Having concluded the model selection process for the regression method, it can be concluded that significant models have been found for all nine stocks of our portfolio. Furthermore the regression models acquired appear to provide sufficient explanatory power in terms of high R^2 values and low error variances of the regression. Also the beta estimates acquired seem to be stable as well. The betas and regression SE's computed for the nine stocks will be used as inputs for the computation of the efficient frontier for the single index regression model. The variance of JSE and the mean log returns of the nine stocks will also be required as inputs for the computation of the regression method's mean-variance frontier.

3.5 Empirical Study: Multiple Index

3.5.1 The data

All of the empirical studies relate to financial stock market data. For this empirical analysis, the data used were the monthly log returns for the stocks Anglos, Jdgroup, Pick and Pay, Remgro, SA-Eagle, Sappi, Sasol, Tigerbrands and Tongaat. The modelling was done over the period July 1988-February 2005. Since the focus of this study is on the multiple index model, we have a basket of eight explanatory variables (indices) to choose from for each of our nine stocks, namely: Anglo Gold, DJ Tran, FT Gold, Gold R, Implats, JSE all share, Paladium and Richemont. These indices will be represented as $X_1, X_2, X_3, X_4, X_5, X_6, X_7$ and X_9 respectively. The data for these 8 indices also spans over the same period. It should be noted to the reader that the data pertaining to this study will also be applicable to the forthcoming multiple index empirical studies performed in the remaining other chapters. Continuous reference will be made to this data as the Multiple Index Data Set (MIDS)

3.5.2 Study Objectives

The purpose of the study in this section is to build multiple index regression models for the stocks $R_1, R_2, R_3, R_4, R_5, R_6, R_7, R_8$ and R_9 using the aforementioned basket of indices defined in the 'The data' sub-section. The objective in trying to find the "best" equation will be to find the simplest model that adequately fits the data. Overall our objective will be to find regression equations with a common set of least number of variables that still explain a percentage of variance in the dependent variables that is comparable to the percentage

explained with all the variables in the equation. We aim to find a common subset of indices that will be significant for all nine stocks in our portfolio.

In the subsequent time series and state space chapters, we expand on these obtained regression models in order for the purpose of constructing the mean-variance frontiers in a multiple index setting.

3.5.3 Methodology

The response variables Anglos, Jdgroup, Pick and Pay, Remgro, SA-Eagle, Sappi, Sasol, Tigerbrands and Tongaat (represented as $R_1, R_2, R_3, R_4, R_5, R_6, R_7, R_8$ and R_9 respectively). The explanatory variables concerned are Anglo Gold, DJ Tran, FT Gold, Gold R, Implats, JSE all share, Paladium and Richemont which we shall represent by $X_1, X_2, X_3, X_4, X_5, X_6, X_7$ and X_8 respectively. The programming language utilised to employ the computations and model building processes is the Time Series Package "EViews 5".

The model building method adopted in this study for all four responses is the Backward Elimination process. For each response all eight explanatory variables will be fitted, (i.e. the complete model), then a process of elimination will commence by deleting those variables that exhibit an insignificant p -value (a small t -statistic i.e. $t < 1.5$ or p -value > 0.05). This rule will not be adhered to in the strictest sense. Variables that demonstrate p -values in the range of 0.05 to approximately 0.12 will be regarded as an acceptable range of significance (i.e. 5% – 12%).

This process is repeated recursively until the only variables remaining are those of a significant p -value. This process is observed for all 9 responses. Ultimately this will

result in nine models with a common subset of indices that explain a larger percentage of variation in the dependent variables opposed to the percentage explained with all 8 indices in the equation.

The principal criteria involved in deciding on these models are the R^2 , $Adj - R^2$, AIC , BIC , the F -statistic and it's corresponding Probability/ P -value. The following subsection discloses the results realised from adopting the backward elimination process and the application of the principal selection criteria.

3.5.4 Primary Findings

The Initial Step

Implementing the methodology outlined in the preceding section, complete models of the eight indices were fit to the monthly log returns data of R_1 , R_2 , R_3 , R_4 , R_5 , R_6 , R_7 , R_8 and R_9 . As the Backward elimination process algorithm suggests, we initially commence by fitting the complete model for all 9 response variables, i.e. fitting all 8 indices to each of the 9 responses (the complete model).

Multiple Regression Models	Model Selection Criteria						
	Adj- R^2	S.E. of reg	D-W stat	AIC	BIC	F-stat	Prob(F-stat)
Model for R1 (Model 1)	0.7208	0.0527	2.0674	-3.0056	-2.8566	64.8847	0.0000
Model for R2 (Model 2)	0.2110	0.1218	2.2356	-1.3293	-1.1804	7.6200	0.0000
Model for R3 (Model 3)	0.2775	0.0867	2.5358	-2.0085	-1.8595	10.5048	0.0000
Model for R4 (Model 4)	0.4438	0.0592	2.2961	-2.7709	-2.6219	20.7451	0.0000
Model for R5 (Model 5)	0.0673	0.0809	2.0284	-2.1460	-1.9971	2.7851	0.0062
Model for R6 (Model 6)	0.3728	0.0877	1.8884	-1.9857	-1.8368	15.7099	0.0000
Model for R7 (Model 7)	0.3816	0.0776	1.8637	-2.2309	-2.0820	16.2721	0.0000
Model for R8 (Model 8)	0.3243	0.0642	2.2130	-2.6099	-2.4609	12.8765	0.0000
Model for R9 (Model 9)	0.2527	0.0791	2.0451	-2.1924	-2.0434	9.3703	0.0000

Table 3.3: Summary of Results: the 9 multiple index regression models

Table 3.3 above represents a summary of the model selection criteria stats for the complete models for all 9 response variables. it is immediately clear that all nine regression models

fitted are exceptionally significant at the 5% level of significance. Secondly it is also observed that Anglos has the highest R^2 , 72.08% followed by Remgro with 44.38% right up to SA-Eagle which exhibits the lowest of the nine.

From above, it can be concluded that the 8 indices explains most of the variation in all nine shares except for SA-Eagle, where it only explains a mere 6.73%. Compared with the single index setting, it is clear that fitting another 7 extra indices as explanatories, has increased the R^2 stat for all 9 responses.

Furthermore, the Durbin-Watson (D-W) statistic is especially greater than 2 for R_1 , R_2 , R_3 , R_4 , R_5 , R_8 and R_9 . This means that there might be a strong presence of negative autocorrelation present. As a consequence time series components will be required to be investigated in the subsequent chapter for these stocks. The converse is true for the remaining two.

Nearly all nine stocks have relatively stable standard error (SE) of regression except for R_2 , which is the only stock that exhibits quite a large SE when compared with the rest. The reader is referred to Appendix 1A for the summary statistics of the beta coefficients of the 8 indices for all nine complete models. Examining the significance of the beta coefficients very carefully, it is clear that most of the time variables X_6 and X_7 are highly significant for nearly all 9 models. The closest variable after them in terms of significance is X_5 . A subset selection procedure based on highest R^2 , $Adj - R^2$, low C_p values resulted in choice of (X_5, X_6, X_7) . indices as explanatory variables in the Multiple index models. It is therefore established that a significant subset of indices to use is (X_5, X_6, X_7) . Thus, the

Multiple Index Regression model that is to be fitted next to our portfolio of 9 responses is

$$R_i = \beta_0 + \beta_1 * Implats + \beta_2 * JSEALS + \beta_3 * Paladium + \varepsilon_i$$

for $i = 1, \dots, 9$.

Table 3.4 below highlights the selection criteria obtained from fitting the new Multiple Index Regression Models for all 9 responses:

Multiple Regression Models	Model Selection Criteria						
	Adj-R ²	S.E. of reg	D-W stat	AIC	BIC	F-stat	Prob(F-stat)
Model for R1 (Model 1)	0.7061	0.0540	2.0318	-2.9787	-2.9125	159.5874	0.0000
Model for R2 (Model 2)	0.2111	0.1218	2.1906	-1.3537	-1.2875	18.6621	0.0000
Model for R3 (Model 3)	0.2482	0.0884	2.5517	-1.9930	-1.9268	22.7913	0.0000
Model for R4 (Model 4)	0.4068	0.0612	2.2866	-2.7308	-2.6646	46.2554	0.0000
Model for R5 (Model 5)	0.0554	0.0815	2.0528	-2.1577	-2.0915	4.8720	0.0027
Model for R6 (Model 6)	0.3774	0.0874	1.9198	-2.0174	-1.9512	41.0095	0.0000
Model for R7 (Model 7)	0.3798	0.0777	1.8304	-2.2523	-2.1861	41.4142	0.0000
Model for R8 (Model 8)	0.3048	0.0651	2.1224	-2.6058	-2.5396	29.9351	0.0000
Model for R9 (Model 9)	0.2358	0.0800	2.0224	-2.1943	-2.1281	21.3642	0.0000

Table 3.4: Summary of model Selection Criteria for new Multiple Index Model

The first thing that one should notice is that all nine $Adj-R^2$'s have decreased very slightly. This is obvious since we have excluded 5 explanatory variables from our initial model, thus we have less explaining power than before. Secondly that all 9 F-statistics are even higher than before, thus we have more significant models than before for all 9 cases. The SE of regression are worse, while the AIC and BIC have worsened slightly as well. Overall, it is clear that we have a much simpler model, whilst still retaining much of the explanatory power with the penalty of incurring slightly higher SE's of regression.

Next we analyse the significance of the beta coefficients, where the reader is referred to appendix 1B. In most cases it was found that at least 2 of the 3 indices were significant, whilst most of the time the 3rd variable proved to be highly insignificant, but still producing a non-zero beta coefficient. Since a subset of at least 3 indices was required we were forced to include X_5 as our 3rd index.

3.5.5 Conclusions

Having concluded the elimination process, the newly found models are presented below in their linear regression forms:

Multiple Reg Models	Definition Of Multiple Reg Structure
Model for R1 (Model 1)	Anglos = $\text{const} + \beta_1 \text{'Implats} + \beta_2 \text{'JSE} + \beta_3 \text{'Palam} + e_1$
Model for R2 (Model 2)	Jdgroup = $\text{const} + \beta_1 \text{'Implats} + \beta_2 \text{'JSE} + \beta_3 \text{'Palam} + e_2$
Model for R3 (Model 3)	Pick and Pay= $\text{const} + \beta_1 \text{'Implats} + \beta_2 \text{'JSE} + \beta_3 \text{'Palam} + e_3$
Model for R4 (Model 4)	Remgro= $\text{const} + \beta_1 \text{'Implats} + \beta_2 \text{'JSE} + \beta_3 \text{'Palam} + e_4$
Model for R5 (Model 5)	SA-Eagle= $\text{const} + \beta_1 \text{'Implats} + \beta_2 \text{'JSE} + \beta_3 \text{'Palam} + e_5$
Model for R6 (Model 6)	Sappi= $\text{const} + \beta_1 \text{'Implats} + \beta_2 \text{'JSE} + \beta_3 \text{'Palam} + e_6$
Model for R7 (Model 7)	Sasol= $\text{const} + \beta_1 \text{'Implats} + \beta_2 \text{'JSE} + \beta_3 \text{'Palam} + e_7$
Model for R8 (Model 8)	Tigerbrands= $\text{const} + \beta_1 \text{'Implats} + \beta_2 \text{'JSE} + \beta_3 \text{'Palam} + e_8$
Model for R9 (Model 9)	Tongaat= $\text{const} + \beta_1 \text{'Implats} + \beta_2 \text{'JSE} + \beta_3 \text{'Palam} + e_9$

Table 3.5: Final Multiple Index Regression Models

where e_i is normally distributed so that $e_i \sim N(0, \sigma^2)$ for $i = 1, \dots, 9$. Also e_t and e_s are independently distributed for all $t \neq s$, so that $Cov(e_t, e_s) = 0$. The error term e_t is a random variable with mean zero, i.e. $E[e_t] = 0$. The above simplified multiple index regression models are to be extended in time series and state space forms in the multiple index empirical studies of subsequent chapters. The multiple regression models established here shall be used to construct the relevant mean-variance frontier later on for comparison purposes.

3.6 List Of Appendices

3.6.1 Appendix 1A: The Complete Models: Summary of Beta coefficients

Variable	Coefficient	Std. Error	t-Statistic	Prob.
C	-0.000589	0.003866	-0.152423	0.8790
X1	0.101782	0.044409	2.291910	0.0230
X2	0.083774	0.060246	1.390525	0.1660
X3	0.035325	0.046673	0.756865	0.4501
X4	0.042498	0.065096	0.499411	0.6181
X5	0.099177	0.037906	2.616425	0.0096
X6	1.132629	0.100587	11.26016	0.0000
X7	0.069716	0.035928	1.940405	0.0538
X8	-0.077823	0.063698	-1.221743	0.2233

R1-Model

Variable	Coefficient	Std. Error	t-Statistic	Prob.
C	0.006061	0.006365	0.952342	0.3421
X1	0.006345	0.073112	0.086778	0.9309
X2	0.104771	0.099185	1.056325	0.2922
X3	-0.184873	0.076839	-2.405985	0.0171
X4	-0.123804	0.140096	-0.883711	0.3780
X5	-0.245012	0.062405	-3.926161	0.0001
X6	1.200789	0.165600	7.251163	0.0000
X7	0.044936	0.059150	0.759694	0.4484
X8	-0.140022	0.104868	-1.335217	0.1834

R3-Model

Variable	Coefficient	Std. Error	t-Statistic	Prob.
C	0.012439	0.005942	2.093445	0.0376
X1	0.106225	0.068253	1.556343	0.1213
X2	-0.134505	0.092593	-1.452653	0.1480
X3	0.036085	0.071732	0.503054	0.6155
X4	-0.135393	0.130785	-1.035235	0.3019
X5	-0.142241	0.058257	-2.441598	0.0155
X6	0.377876	0.154593	2.444329	0.0154
X7	0.020957	0.055219	0.379527	0.7047
X8	0.021254	0.097898	0.217099	0.8284

R5-Model

Variable	Coefficient	Std. Error	t-Statistic	Prob.
C	0.005636	0.005695	0.989684	0.3236
X1	0.026083	0.065417	0.398720	0.6905
X2	-0.141481	0.088745	-1.594236	0.1125
X3	-0.017314	0.068751	-0.251831	0.8014
X4	0.153618	0.125350	1.225511	0.2219
X5	-0.077377	0.055837	-1.385786	0.1674
X6	1.059186	0.148169	7.148494	0.0000
X7	0.126334	0.052924	2.387080	0.0180
X8	-0.007360	0.093830	-0.078436	0.9376

R7-Model

Variable	Coefficient	Std. Error	t-Statistic	Prob.
C	-0.000273	0.005806	-0.047107	0.9625
X1	-0.062528	0.066689	-0.937607	0.3496
X2	0.104560	0.090471	1.155731	0.2492
X3	-0.095300	0.070088	-1.359714	0.1755
X4	0.097019	0.127788	0.759219	0.4487
X5	0.028600	0.056922	0.502442	0.6159
X6	0.870227	0.151051	5.761158	0.0000
X7	0.115886	0.053953	2.147891	0.0330
X8	-0.189693	0.095655	-1.983101	0.0488

R9-Model

Variable	Coefficient	Std. Error	t-Statistic	Prob.
C	0.005490	0.008938	0.614234	0.5398
X1	-0.090778	0.102673	-0.884146	0.3777
X2	0.053136	0.139288	0.381483	0.7033
X3	0.049342	0.107907	0.457266	0.6480
X4	-0.268062	0.196741	-1.362515	0.1746
X5	-0.135206	0.087637	-1.542794	0.1245
X6	1.429260	0.232556	6.145883	0.0000
X7	-0.052600	0.083066	-0.633231	0.5273
X8	-0.170356	0.147269	-1.156766	0.2488

R2-Model

Variable	Coefficient	Std. Error	t-Statistic	Prob.
C	0.007617	0.004347	1.752080	0.0814
X1	0.011564	0.049939	0.231558	0.8171
X2	-0.091349	0.067747	-1.348379	0.1791
X3	-0.007367	0.052484	-0.140366	0.8885
X4	-0.221345	0.095691	-2.313112	0.0218
X5	-0.056897	0.042625	-1.334824	0.1835
X6	0.795218	0.113111	7.030404	0.0000
X7	-0.008861	0.040402	-0.219332	0.8266
X8	0.244193	0.071629	3.409124	0.0008

R4-Model

Variable	Coefficient	Std. Error	t-Statistic	Prob.
C	-0.004382	0.006438	-0.680639	0.4969
X1	-0.030853	0.073948	-0.417223	0.6770
X2	-0.093354	0.100319	-0.930570	0.3533
X3	-0.064112	0.077718	-0.824926	0.4104
X4	-0.106857	0.141699	-0.754113	0.4517
X5	0.019913	0.063119	0.315487	0.7527
X6	1.117483	0.167494	6.671787	0.0000
X7	0.235016	0.059827	3.928290	0.0001
X8	-0.080177	0.106068	-0.755903	0.4506

R6-Model

Variable	Coefficient	Std. Error	t-Statistic	Prob.
C	0.004861	0.004712	1.031649	0.3035
X1	-0.039877	0.054124	-0.736760	0.4622
X2	0.041954	0.073426	0.571382	0.5684
X3	0.007588	0.056883	0.133395	0.8940
X4	-0.259538	0.103712	-2.502496	0.0132
X5	-0.111051	0.046198	-2.403806	0.0172
X6	0.960030	0.122592	7.831109	0.0000
X7	-0.034925	0.043788	-0.797584	0.4261
X8	-0.087256	0.077633	-1.123951	0.2625

R8-Model

3.6.2 Appendix 1B: The Final Models: Summary of Beta coefficients

Variable	Coefficient	Std. Error	t-Statistic	Prob.
C	-0.002278	0.003893	-0.585275	0.5590
X5	0.131949	0.036450	3.620050	0.0004
X6	1.207564	0.080458	15.00861	0.0000
X7	0.051472	0.035794	1.437992	0.1520

R1-Final Model

Variable	Coefficient	Std. Error	t-Statistic	Prob.
C	0.007335	0.006372	1.151079	0.2511
X5	-0.307985	0.059667	-5.161740	0.0000
X6	1.021138	0.131708	7.753055	0.0000
X7	0.071212	0.058594	1.215350	0.2257

R3-Final Model

Variable	Coefficient	Std. Error	t-Statistic	Prob.
C	0.009710	0.005869	1.654546	0.0996
X5	-0.114772	0.054953	-2.088559	0.0380
X6	0.459024	0.121301	3.784160	0.0002
X7	-0.012172	0.053964	-0.225565	0.8218

R5-Final Model

Variable	Coefficient	Std. Error	t-Statistic	Prob.
C	0.005252	0.005598	0.938197	0.3493
X5	-0.051944	0.052414	-0.991039	0.3229
X6	1.012362	0.115698	8.750051	0.0000
X7	0.122534	0.051471	2.380623	0.0182

R7-Final Model

Variable	Coefficient	Std. Error	t-Statistic	Prob.
C	0.001350	0.005762	0.234298	0.8150
X5	-0.011782	0.053956	-0.218366	0.8274
X6	0.655586	0.119102	5.504392	0.0000
X7	0.148862	0.052986	2.809467	0.0055

R9-Final Model

Variable	Coefficient	Std. Error	t-Statistic	Prob.
C	0.004717	0.008772	0.537719	0.5914
X5	-0.190363	0.082141	-2.317520	0.0215
X6	1.294221	0.181317	7.137908	0.0000
X7	-0.045490	0.080664	-0.563949	0.5734

R2-Final Model

Variable	Coefficient	Std. Error	t-Statistic	Prob.
C	0.007789	0.004406	1.767547	0.0787
X5	-0.063865	0.041261	-1.547821	0.1233
X6	0.947516	0.091079	10.40321	0.0000
X7	-0.021930	0.040519	-0.541231	0.5890

R4-Final Model

Variable	Coefficient	Std. Error	t-Statistic	Prob.
C	-0.004430	0.006295	-0.703691	0.4825
X5	-0.014977	0.058945	-0.254081	0.7997
X6	0.965444	0.130113	7.420021	0.0000
X7	0.243997	0.057885	4.215231	0.0000

R6-Final Model

Variable	Coefficient	Std. Error	t-Statistic	Prob.
C	0.004271	0.004691	0.910480	0.3637
X5	-0.155539	0.043922	-3.541241	0.0005
X6	0.892107	0.096953	9.201410	0.0000
X7	-0.032662	0.043132	-0.757257	0.4498

R8-Final Model

Chapter 4

Linear Time Series Models and Empirical Studies

4.1 Introduction

Why do we concentrate exclusively on time series techniques when, for example, cross sectional modelling plays an important role in empirical investigations of the Capital Asset Pricing Model (CAPM). As an influential and early example the reader is referred to Fama and MacBeth (1973)? The answer is that , apart from the usual considerations of personal expertise and interest, it is because time series analysis, in both its theoretical and empirical aspects, has been for many years an integral part of the study of financial markets, with empirical research commencing with papers by Working (1934), Cowles (1933, 1944) and Cowles and Jones (1937).

Working focused attention on a previously noted characteristic of commodity and stock prices: namely, that they resemble cumulations of purely random changes. Cowles investigated the ability of market analysts and financial services to predict future price changes, finding that there was little evidence that they could. Cowles and Jones reported evidence of positive correlation between successive price changes but, as Cowles (1960) was later to remark, this probably due to their taking monthly averages of daily or weekly prices before computing changes.

The predictability of price changes has since become a major theme of financial research but, surprisingly, little more was published until Kendall's (1953) study, in which he found that the weekly changes in a wide variety of financial prices could not be predicted from either past changes in the series or from past changes in other price series. This seems to have been the first explicit reporting of this oft-quoted property of financial prices, although further impetus research on price predictability was only provided by the publication of the papers by Roberts (1959) and Osbourne (1959). The former presents a largely heuristic argument for why successive price changes should be independent, while the latter develops the proposition that it is not absolute price changes but the logarithmic price changes which are independent of each other.

The stimulation provided by these papers was such that numerous articles appeared over the next few years investigating the hypothesis that price changes are independent, a hypothesis that came to be termed the random walk model, in recognition of the similarity of the evolution of a price series to the random stagger of a drunk man. The most formal way to state formally the random walk model is as

$$P_t = P_{t-1} + a_t \quad (1)$$

where P_t is the price observed at the beginning of time t and a_t is the error term which has zero mean and whose values are independent of each other. The price change, $\Delta P_t = P_t - P_{t-1}$, is thus simply a_t and hence is independent of past price changes. Note that, by successive backward substitution in (1), we can write the current price as the cumulation

of all past errors, i.e.

$$P_t = \sum_{i=1}^t a_i$$

so that the random walk model implies that prices are indeed generated by Working's cumulation of purely random changes.

This chapter is organised as follows: In the subsequent section we introduce Autoregressive Models in the time series context, where we present the estimation, identification and definitions of such models. Again their estimation, identification and important definitions regarding these models are introduced. Examples are provided in most of the literature sections of this chapter in order to familiarise the reader with the concepts and models regarding time series. Finally the last section presents an empirical study involving the literature discussed in this chapter. (Mills(1999))

4.2 Objectives of Time Series analysis

Time series analysis typically involve setting up a hypothetical probability model to enhance our understanding of the data. Many of the ideas such as model selection, parameter estimation and goodness of fit checks, extend to time series analysis. Once we have identified a satisfactory model, it can be used in a variety of ways depending on the application. There are several possible objectives in analysing a time series. These objectives may be classified as:

- 1. Description:** When presented with a time series, the first step in the analysis is usually to plot the observations against time to give what is called a time plot, and then to obtain simple descriptive measures of the main properties of the series. The time plot is a

powerful descriptive tool which can illustrate graphical effects like seasonality and depict the trend type (i.e. multiplicative or linear etc.). One is asking for trouble when trying to analyse a time series without plotting it first. not only does the plot help illustrate seasonality and trend, but it will also reveal any 'wild' observations or outliers that do not appear to be consistent with the data. Other features to look for in a time plot include sudden or gradual changes in the properties of the series. For example, a step change in the level of the series would be very important to notice if it existed. Also any changes in the seasonal pattern should be noted. Turning points and discontinuities in the series should be checked for, then different models may need to be fitted to different parts of the series. Figure 4.1 below illustrates the time plot for the stock Anglos, which clearly shows that there is a multiplicative trend, hence the prices are increasing in an exponential manner. It also confirms that there is no seasonality present, otherwise we would become millionaires.

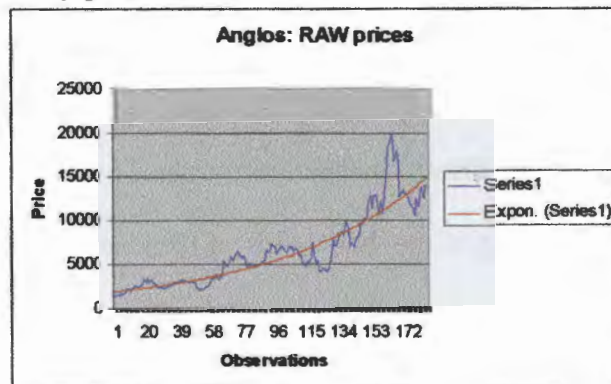


Figure 4.1: Time Plot of Raw prices for ANGLOS

2. Modelling: When observations are taken on two or more variables, it may be possible to use the variation in one time series to explain the variation in another series. Such a technique may lead to deeper understanding of the mechanism that generated a

given time series. Occasionally multiple regression models are helpful here but they are not designed to handle time series data with all the correlations inherent therein, hence another class of models will be considered.

3. Forecasting: Given an observed time series, it is usually the case that one wishes to predict the forthcoming values of the series. This is especially typical in the analysis of economic and financial data. In our case the focus is on stock exchange data.

4. Control: Time series are sometimes collected and analysed so as to improve control over some physical or economic system. Control problems are closely associated with forecasting in some sense. For instance if one was to accurately predict some unfortunate circumstance, appropriate measure can be put into place prior to the unfavourable action occurring. Control procedures vary considerably in style and sophistication. In statistical quality control, the observations are plotted on control charts and the controller takes action as a result of studying the charts.

With regards to this dissertation, we will only consider objectives one to three, since four is out of the scope of the current study.

4.3 Autoregressive Models

Autoregressive series are important because:

1. They have a natural interpretation the next value observed is a slight perturbation of the most recent observation.
2. It is easy to estimate their parameters. It can be done with standard regression software.

3. They are easy to forecast. Again standard regression software will do the job.

The First-Order Autoregressive Model.

The $AR(1)$ process is given by

$$r_t = \phi_0 + \phi_1 r_{t-1} + e_t \quad (1)$$

with e_t assumed to be white noise with

$$E(e_t) = 0 \text{ and } var(e_t) = \sigma^2.$$

Assuming that the process is weakly stationary, we have

$$E(r_t) = \mu, \text{ var}(r_t) = \gamma_0 \text{ and}$$

$$cov(r_t r_{t-j}) = \gamma_j$$

where μ and γ_0 are constant and γ_j is a function of j and not t . Since $E(e_t) = 0$, then the mean is,

$$E(r_t) = \phi_0 + \phi_1 E(r_{t-1})$$

Under the stationarity condition $E(r_t) = E(r_{t-1}) = \mu$ and hence

$$\mu = \phi_0 + \phi_1 \mu \text{ or } E(r_t) = \mu = \frac{\phi_0}{1 - \phi_1}.$$

This result has two implications. First the mean of r_t exists if $\phi_1 \neq 1$. Second, the mean of r_t is zero if and only if $\phi_0 = 0$. Thus for a stationary $AR(1)$ process, the constant term ϕ_0 is related to the mean of r_t and $\phi_0 = 0$ implies $E(r_t) = 0$. Next using $\phi_0 = (1 - \phi_1)\mu$ the $AR(1)$ model can be rewritten as

$$r_t - \mu = \phi_1 (r_{t-1} - \mu) + e_t. \quad (2)$$

By repeated substitutions we get

$$\begin{aligned} r_t - \mu &= e_t + \phi_1 e_{t-1} + \phi_1^2 e_{t-2} + \dots \quad (3) \\ &= \sum_{i=0}^{\infty} \phi_1^i e_{t-i}. \end{aligned}$$

Thus $r_t - \mu$ is a linear function of e_{t-i} for $i \geq 0$. Using this property and the independence of the series $\{e_t\}$, we obtain $E(r_t - \mu)e_{t+1} = 0$. By the stationarity assumption

$$\text{cov}(r_{t-1}, e_t) = E[(r_{t-1} - \mu)e_t] = 0.$$

This result can also be seen from the result that r_{t-1} occurred before time t and e_t does not depend on any past information. Taking the square, then the expectation of (2), we obtain

$$\text{var}(r_t) = \phi_1^2 \text{var}(r_{t-1}) + \sigma^2, \text{ since } \text{cov}(r_{t-1}, e_t) = 0.$$

Thus, under stationarity $\text{var}(r_t) = \text{var}(r_{t-1})$ so that (note that $\text{var}(e_t) = \sigma^2$)

$$\text{var}(r_t) = \frac{\sigma^2}{1 - \phi_1^2}$$

provided that $\phi_1^2 < 1$. The requirement of $\phi_1^2 < 1$ results from the fact that the variance of the random variable r_t is bounded and non-negative. Consequently the weak stationarity of an $AR(1)$ process implies that $-1 < \phi_1 < 1$. But, if $-1 < \phi_1 < 1$ then by (2) and the independence of e_t , we can show that the mean and variance of r_t are finite. In addition, by the Cauchy-Schwartz inequality, all the covariances of r_t are finite. Therefore, the $AR(1)$ model is weakly stationary. Thus, in summary, the necessary and sufficient condition for the $AR(1)$ process to be stationary is $|\phi_1| < 1$.

Autocorrelation Function of an $AR(1)$ Model.

Multiplying equation (2) above by e_t , using the independence between e_t and r_{t-1} and taking expectation, we obtain

$$\begin{aligned} E[e_t(r_t - \mu)] &= E[e_t(r_{t-1} - \mu)] + E(e_t^2) \\ &= E(e_t^2) = \sigma^2. \end{aligned}$$

Thus multiplying (2) by $(r_{t-l} - \mu)$ and taking expectations, we have

$$\begin{aligned} E[(r_t - \mu)(r_{t-l} - \mu)] &= E[\phi_1(r_{t-1} - \mu)(r_{t-l} - \mu) + E[e_t(r_{t-l} - \mu)]] \\ &= \phi_1 \gamma_l + \sigma^2 \text{ for } l = 0 \\ &= \phi_1 \gamma_{l-1} \text{ for } l > 0, \end{aligned}$$

where we use $\gamma_l = \gamma_{-l}$. Consequently, for a weak stationary $AR(1)$ model in equation (2)

$$\text{var}(r_t) = \gamma_0, \text{ and } \gamma_l = \phi_1 \gamma_{l-1} \text{ for } l > 0.$$

Thus the ACF of ρ_l satisfies

$$\rho_l = \phi_1 \rho_{l-1} \text{ for } l \geq 0.$$

Because $\rho_0 = 1$ we have

$$\rho_l = \phi_1^l, \tag{4}$$

thus the ACF of a weakly stationary $AR(1)$ series decays exponentially with rate ϕ_1 and starting at $\rho_0 = 1$. Alternative Derivation of the ACF for the $AR(1)$ Model. Now since from (3)

$$r_t - \mu = e_t + \phi_1 e_{t-1} + \phi_1^2 e_{t-2} + \phi_1^3 e_{t-3} + \dots$$

The variance is

$$\begin{aligned}
 \gamma_0 &= E(r_t - \mu)^2 \\
 &= E(e_t + \phi_1 e_{t-1} + \phi_1^2 e_{t-2} + \phi_1^3 e_{t-3} + \dots)^2 \\
 &= (1 + \phi_1^2 + \phi_1^4 + \phi_1^6 + \dots) \sigma^2 \\
 &= \sigma^2 / (1 - \phi_1^2),
 \end{aligned}$$

while the l th autocovariance is

$$\begin{aligned}
 \gamma_l &= E(r_t - \mu)(r_{t-l} - \mu) \\
 &= E(e_t + \phi_1 e_{t-1} + \phi_1^2 e_{t-2} + \phi_1^3 e_{t-3} + \dots) \\
 &\quad \times (e_{t-l} + \phi_1 e_{t-l-1} + \phi_1^2 e_{t-l-2} + \phi_1^3 e_{t-l-3} + \dots) \\
 &= [\phi_1^l + \phi_1^{l+2} + \phi_1^{l+4} + \dots] \sigma^2 \\
 &= \phi_1^l [1 + \phi_1^2 + \phi_1^4 + \dots] \sigma^2 \\
 &= [\phi_1^l / (1 - \phi_1^2)] \sigma^2.
 \end{aligned}$$

The autocorrelation function is therefore

$$\begin{aligned}
 \rho_j &= \frac{\gamma_j}{\gamma_0}, \quad \rho_0 = 1 \\
 &= \phi_1^j.
 \end{aligned} \tag{5}$$

By using the Lag operator $Lr_t = r_{t-1}$ and $L^q r_t = r_{t-q}$ we get from (2)

$$\begin{aligned}
 (1 - \phi_1 L)(r_t - \mu) &= e_t \text{ or} \\
 \Phi(L)(r_t - \mu) &= e_t, \text{ with } \Phi(L) = 1 - \phi_1 L
 \end{aligned} \tag{6}$$

By repeated substitutions we get

$$\begin{aligned} r_t - \mu &= e_t + \phi_1 e_{t-1} + \phi_1^2 e_{t-2} + \dots \quad (3) \\ &= \sum_{i=0}^{\infty} \phi_1^i e_{t-i}. \end{aligned}$$

Thus $r_t - \mu$ is a linear function of e_{t-i} for $i \geq 0$. Using this property and the independence of the series $\{e_t\}$, we obtain $E(r_t - \mu)e_{t+1} = 0$. By the stationarity assumption

$$\text{cov}(r_{t-1}, e_t) = E[(r_{t-1} - \mu)e_t] = 0.$$

This result can also be seen from the result that r_{t-1} occurred before time t and e_t does not depend on any past information. Taking the square, then the expectation of (2), we obtain

$$\text{var}(r_t) = \phi_1^2 \text{var}(r_{t-1}) + \sigma^2, \text{ since } \text{cov}(r_{t-1}, e_t) = 0.$$

Thus, under stationarity $\text{var}(r_t) = \text{var}(r_{t-1})$ so that (note that $\text{var}(e_t) = \sigma^2$)

$$\text{var}(r_t) = \frac{\sigma^2}{1 - \phi_1^2}$$

provided that $\phi_1^2 < 1$. The requirement of $\phi_1^2 < 1$ results from the fact that the variance of the random variable r_t is bounded and non-negative. Consequently the weak stationarity of an $AR(1)$ process implies that $-1 < \phi_1 < 1$. But, if $-1 < \phi_1 < 1$ then by (2) and the independence of e_t , we can show that the mean and variance of r_t are finite. In addition, by the Cauchy-Schwartz inequality, all the covariances of r_t are finite. Therefore, the $AR(1)$ model is weakly stationary. Thus, in summary, the necessary and sufficient condition for the $AR(1)$ process to be stationary is $|\phi_1| < 1$.

Autocorrelation Function of an $AR(1)$ Model.

Multiplying equation (2) above by e_t , using the independence between e_t and r_{t-1} and taking expectation, we obtain

$$\begin{aligned} E[e_t(r_t - \mu)] &= E[e_t(r_{t-1} - \mu)] + E(e_t^2) \\ &= E(e_t^2) = \sigma^2. \end{aligned}$$

Thus multiplying (2) by $(r_{t-l} - \mu)$ and taking expectations, we have

$$\begin{aligned} E[(r_t - \mu)(r_{t-l} - \mu)] &= E[\phi_1(r_{t-1} - \mu)(r_{t-l} - \mu) + E[e_t(r_{t-l} - \mu)]] \\ &= \phi_1\gamma_l + \sigma^2 \text{ for } l = 0 \\ &= \phi_1\gamma_{l-1} \text{ for } l > 0, \end{aligned}$$

where we use $\gamma_l = \gamma_{-l}$. Consequently, for a weak stationary $AR(1)$ model in equation (2)

$$\text{var}(r_t) = \gamma_0, \text{ and } \gamma_l = \phi_1\gamma_{l-1} \text{ for } l > 0.$$

Thus the ACF of ρ_l satisfies

$$\rho_l = \phi_1\rho_{l-1} \text{ for } l \geq 0.$$

Because $\rho_0 = 1$ we have

$$\rho_l = \phi_1^l, \tag{4}$$

thus the ACF of a weakly stationary $AR(1)$ series decays exponentially with rate ϕ_1 and starting at $\rho_0 = 1$. Alternative Derivation of the ACF for the $AR(1)$ Model. Now since from (3)

$$r_t - \mu = e_t + \phi_1 e_{t-1} + \phi_1^2 e_{t-2} + \phi_1^3 e_{t-3} + \dots$$

The variance is

$$\begin{aligned}
 \gamma_0 &= E(r_t - \mu)^2 \\
 &= E(e_t + \phi_1 e_{t-1} + \phi_1^2 e_{t-2} + \phi_1^3 e_{t-3} + \dots)^2 \\
 &= (1 + \phi_1^2 + \phi_1^4 + \phi_1^6 + \dots) \sigma^2 \\
 &= \sigma^2 / (1 - \phi_1^2),
 \end{aligned}$$

while the l th autocovariance is

$$\begin{aligned}
 \gamma_l &= E(r_t - \mu)(r_{t-l} - \mu) \\
 &= E(e_t + \phi_1 e_{t-1} + \phi_1^2 e_{t-2} + \phi_1^3 e_{t-3} + \dots) \\
 &\quad \times (e_{t-l} + \phi_1 e_{t-l-1} + \phi_1^2 e_{t-l-2} + \phi_1^3 e_{t-l-3} + \dots) \\
 &= [\phi_1^l + \phi_1^{l+2} + \phi_1^{l+4} + \dots] \sigma^2 \\
 &= \phi_1^l [1 + \phi_1^2 + \phi_1^4 + \dots] \sigma^2 \\
 &= [\phi_1^l / (1 - \phi_1^2)] \sigma^2.
 \end{aligned}$$

The autocorrelation function is therefore

$$\begin{aligned}
 \rho_j &= \frac{\gamma_j}{\gamma_0}, \quad \rho_0 = 1 \\
 &= \phi_1^j.
 \end{aligned} \tag{5}$$

By using the Lag operator $Lr_t = r_{t-1}$ and $L^q r_t = r_{t-q}$ we get from (2)

$$\begin{aligned}
 (1 - \phi_1 L)(r_t - \mu) &= e_t \text{ or} \\
 \Phi(L)(r_t - \mu) &= e_t, \text{ with } \Phi(L) = 1 - \phi_1 L
 \end{aligned} \tag{6}$$

Let $x = 1/L$ then we get the characteristic equation $\Phi(\frac{1}{x}) = 0$ of the $AR(1)$ model. The characteristic equation plays an important role in identifying Linear Time Series Processes.

The characteristic equation is

$$\Phi\left(\frac{1}{x}\right) = x - \phi_1 = 0 \quad (7)$$

and the only solution of the characteristic equation is

$$x = \phi_1. \quad (8)$$

But for the process to be stationary $|\phi_1| < 1$. Thus the solution of the characteristic equation requires $|\phi_1| < 1$ for the process to be stationary, or, the solution must be within the unit circle. Note that some writers use as characteristic equation, with $z = L$,

$$\Phi_1(z) = 1 - \phi_1 z = 0 \quad (9)$$

with solution $z = 1/\phi_1$ and $|z| = |1/\phi_1| > 1$, requires $|\phi_1| < 1$ for the process to be stationary. Thus the solutions to the characteristic equation $\Phi_1(z) = 0$, must all lie outside the unit circle.

Examples.

We give below an example. For the $AR(1)$ series

$$r_t = 2 + 0.5r_{t-1} + e_t$$

where e_t is Gaussian white noise. The autocorrelation function (called the Correlogram) is also displayed (below in figure 4.2) of the typical realization of the $AR(1)$ time series process and clearly shows exponential decay. The $AR(1)$ can be identified by a single spike for the partial autocorrelation function. This will be discussed later. The realization of the $AR(1)$ time series process is illustrated below in figure 4.3.

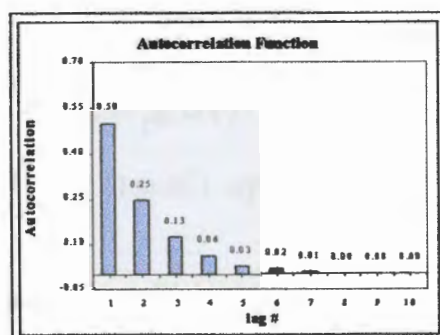
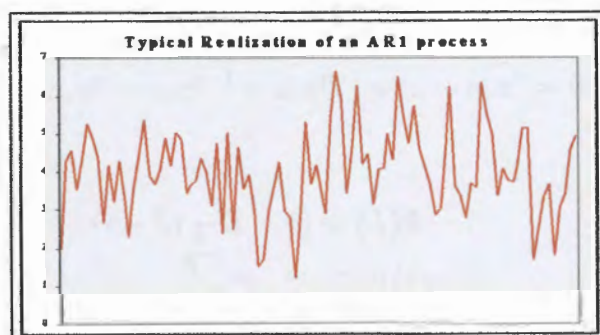


Figure 4.2 : ACF of AR(1)

Figure 4.3: typical realization of the $AR(1)$

The p th-order Autoregressive Model.

The $AR(p)$ model is given by

$$r_t = \phi_0 + \phi_1 r_{t-1} + \dots + \phi_p r_{t-p} + e_t. \quad (10)$$

with $E(e_t) = 0$ and $\text{var}(e_t) = \sigma^2$. Under the stationarity condition $E(r_t) = E(r_{t-j}) = \mu$, so that

$$E(r_t) = \mu = \phi_0 + \phi_1 \mu + \dots + \phi_p \mu$$

or

$$\mu = \frac{\phi_0}{(1 - \phi_1 - \phi_2 - \dots - \phi_p)} \quad (11)$$

provided the denominator is not zero. Using (11) in (10) we get

$$r_t - \mu = \phi_1(r_{t-1} - \mu) + \dots + \phi_p(r_{t-p} - \mu) + e_t. \quad (12)$$

In lag operator form

$$(1 - \phi_1 L - \phi_2 L^2 - \dots - \phi_p L^p)(r_t - \mu) = e_t, \text{ or} \quad (13)$$

$$\Phi(L)(r_t - \mu) = e_t,$$

where

$$\Phi(L) = (1 - L - L^2 - \dots - L^p),$$

and (13) is the difference equation form of the $AR(p)$ model. If $x = 1/L$ then the equation

$$\Phi\left(\frac{1}{x}\right) = x^p - \phi_1 x^{p-1} - \phi_2 x^{p-1} - \dots - \phi_p = 0$$

is the characteristic equation of the $AR(p)$ process. This equation plays an important role in the study of the invertibility and stationarity of an $AR(p)$ process. We now prove a general theorem which will have important applications for the linear and general time series processes that follow later. For convenience sake we assume that $\mu = 0$.

Theorem The Wold Decomposition Theorem. If x_1, x_2, \dots, x_p denote the roots of the p 'th order equation $\Phi\left(\frac{1}{x}\right) = 0$, then the $AR(p)$ model

$$\Phi(L)r_t = e_t$$

can be written as a convergent series in e_t, e_{t-1}, \dots such that

$$\begin{aligned} r_t &= (\Phi(L))^{-1} e_t \\ &= \Psi(L) e_t \\ &= \sum_{r=0}^{\infty} \psi_r e_{t-r}, \quad \psi_0 = 1 \end{aligned} \quad (14)$$

only if $|x_i| < 1$ for all roots $x_i, i = 1, 2, \dots, p$ where

$$\Psi(L) = 1 - \psi_1 L - \psi_2 L^2 - \dots$$

Lemma If x_1, x_2, \dots, x_p are the roots of a p 'th order equation in x such that

$$a_0 x^p + a_1 x^{p-1} + a_2 x^{p-2} + \dots + a_p x^0 = 0$$

then it follows that

$$\begin{aligned} \sum_{i=1}^p x_i &= -a_1/a_0 \\ \sum_{i < j} x_i x_j &= (-1)^2 a_2/a_0 \end{aligned}$$

ect

until

$$\prod_{i=1}^p x_i = (-1)^p a_p/a_0.$$

Proof (Wold Decomposition Theorem)

If x_1, x_2, \dots, x_p are the roots of the equation $\Phi(\frac{1}{x}) = 0$

$$\text{i.e. } x^p - \phi_1 x^{p-1} - \dots - \phi_p x^0 = 0$$

then

$$\prod_{i=1}^p x_i = (-1)^p (-\phi_p) = (-1)^{p+1} \phi_p. \quad (15)$$

So if z_1, z_2, \dots, z_p are the roots of the equation $\Phi(z) = 0$ then $z_i = 1/x_i$ and from (15) follows

$$\prod_{i=1}^p z_i = (-1)^p \left(\frac{-1}{\phi_p} \right). \quad (16)$$

Now

$$\begin{aligned} \Phi(z) &= -\phi_p z^p - \dots - \phi_1 z + 1 \\ &= -\phi_p \left(z^p + \frac{\phi_{p-1}}{\phi_p} z^{p-1} + \dots + \frac{\phi_1}{\phi_p} z - \frac{1}{\phi_p} \right) \\ &= -\phi_p \left(\sum_{r=1}^p \frac{\phi_r}{\phi_p} z^r - \frac{1}{\phi_p} \right). \end{aligned} \quad (17)$$

Since z_1, z_2, \dots, z_p are the roots of the equation $\Phi(z) = 0$ one can factorize the left hand side of $\Phi(z) = 0$ such that

$$\Phi(z) = k \times \prod_{i=1}^p (z - z_i) = 0 \quad (18)$$

where k is a constant. Now

$$\Phi(L) = 1 - \phi_1 L - \phi_2 L^2 - \dots - \phi_p L^p$$

and thus

$$\Phi(z) = 1 - \phi_1 z - \phi_2 z^2 - \dots - \phi_p z^p = 0. \quad (19)$$

Now dividing equation (19) by ϕ_p one has

$$\begin{aligned} \Phi(z)/\phi_p &= (1 - \phi_1 z - \phi_2 z^2 - \dots - \phi_p z^p)/\phi_p = 0/\phi_p = 0 \\ &= \left(\frac{1}{\phi_p} - \frac{\phi_1}{\phi_p} z - \frac{\phi_2}{\phi_p} z^2 - \dots - z^p \right) = 0 \\ &= - \left(\sum_{r=0}^p \frac{\phi_r}{\phi_p} z^r - \frac{1}{\phi_p} \right) = 0 \end{aligned}$$

So if z_1, z_2, \dots, z_p are the roots of the equation $\Phi(z) = 0$ then $z_i = 1/x_i$ and from (15) follows

$$\prod_{i=1}^p z_i = (-1)^p \left(\frac{-1}{\phi_p} \right). \quad (16)$$

Now

$$\begin{aligned} \Phi(z) &= -\phi_p z^p - \dots - \phi_1 z + 1 \\ &= -\phi_p \left(z^p + \frac{\phi_{p-1}}{\phi_p} z^{p-1} + \dots + \frac{\phi_1}{\phi_p} z - \frac{1}{\phi_p} \right) \\ &= -\phi_p \left(\sum_{r=1}^p \frac{\phi_r}{\phi_p} z^r - \frac{1}{\phi_p} \right). \end{aligned} \quad (17)$$

Since z_1, z_2, \dots, z_p are the roots of the equation $\Phi(z) = 0$ one can factorize the left hand side of $\Phi(z) = 0$ such that

$$\Phi(z) = k \times \prod_{i=1}^p (z - z_i) = 0 \quad (18)$$

where k is a constant. Now

$$\Phi(L) = 1 - \phi_1 L - \phi_2 L^2 - \dots - \phi_p L^p$$

and thus

$$\Phi(z) = 1 - \phi_1 z - \phi_2 z^2 - \dots - \phi_p z^p = 0. \quad (19)$$

Now dividing equation (19) by ϕ_p one has

$$\begin{aligned} \Phi(z)/\phi_p &= (1 - \phi_1 z - \phi_2 z^2 - \dots - \phi_p z^p)/\phi_p = 0/\phi_p = 0 \\ &= \left(\frac{1}{\phi_p} - \frac{\phi_1}{\phi_p} z - \frac{\phi_2}{\phi_p} z^2 - \dots - z^p \right) = 0 \\ &= - \left(\sum_{r=0}^p \frac{\phi_r}{\phi_p} z^r - \frac{1}{\phi_p} \right) = 0 \end{aligned}$$

or

$$\Phi(z) = -\phi_p \left(\sum_{r=0}^p \frac{\phi_r}{\phi_p} z^r - \frac{1}{\phi_p} \right) = 0 \quad (20)$$

Equation (20) is now in the form given in equation (18) and from (17) follows $k = -\phi_p$ and

$$\prod_{i=1}^p (z - z_i) = \left(\sum_{r=0}^p \frac{\phi_r}{\phi_p} z^r - \frac{1}{\phi_p} \right). \quad (21)$$

Substituting (21) in (20) one has

$$\Phi(z) = -\phi_p \prod_{i=1}^p (z - z_i). \quad (22)$$

Thus

$$\begin{aligned}
 [\Phi(z)]^{-1} &= \left[-\phi_p \prod_{i=1}^p (z - z_i) \right]^{-1} \\
 &= \left[-\phi_p \prod_{i=1}^p (-z_i) \prod_{i=1}^p \left(1 - \frac{z}{z_i}\right) \right]^{-1} \\
 &= \left[(-1)^{p+1} \phi_p \prod_{i=1}^p z_i \prod_{i=1}^p \left(1 - \frac{z}{z_i}\right) \right]^{-1} \\
 &= \left[(-1)^{p+1} \phi_p (-1)^{p+1} \frac{1}{\phi_p} \prod_{i=1}^p (1 - zx_i) \right]^{-1} \\
 &= \left[\prod_{i=1}^p \left(1 - \frac{z}{z_i}\right) \right]^{-1} \\
 &= \prod_{i=1}^p \left(1 - \frac{z}{z_i}\right)^{-1} \\
 &= \prod_{i=1}^p (1 - zx_i)^{-1} \text{ since } x_i = 1/z_i. \tag{2} \\
 &= \prod_{i=1}^p \sum_{r=0}^{\infty} (zx_i)^r \text{ if } |zx_i| < 1 \tag{2} \\
 &= \sum_{r=0}^{\infty} \left(\sum_{i_1 \leq i_2 \leq \dots \leq i_r} x_{i_1} x_{i_2} \dots x_{i_r} \right) z^r \\
 &= \sum_{r=0}^{\infty} \psi_r z^r
 \end{aligned}$$

which is an infinite power series in z . If $|x_i| < 1$ the power series will converge absolutely

for any such z such that $|z| < \min \left| \frac{1}{x_i} \right|$ for all $i = 1, \dots, p$. Then we can write

$$(\Phi(L))^{-1} = \Psi(L) = \sum_{r=0}^{\infty} \psi_r L^r$$

so that

$$\begin{aligned}
 r_t &= \Psi(L)e_t \\
 &= e_t + \psi_1 e_{t-1} + \psi_2 e_{t-2} + \dots
 \end{aligned}$$

Remarks:

(1) Since

$$\Psi(L) = \prod_{i=1}^p \sum_{r=0}^{\infty} (Lx_i)^r$$

from (24) with $L = z$ we get

$$\begin{aligned} \Psi(L) &= \prod_{i=1}^p \sum_{r=0}^{\infty} (Lx_i)^r \\ &= 1 + (x_1 + x_2 + \cdots + x_p)L \\ &\quad + (x_1^2 + x_2^2 + \cdots + x_p^2 + x_1x_2 + x_1x_3 + \cdots + x_{p-1}x_p)L^2 \\ &\quad + \cdots \\ &= 1 + \psi_1 L + \psi_2 L^2 + \cdots \end{aligned}$$

we see that the ψ_i -weights can be calculated from the roots of the characteristic equation $\Phi(\frac{1}{x}) = 0$ as follows

$$\begin{aligned} \psi_1 &= \sum_{i=1}^p x_i \\ \psi_2 &= \sum_{i \leq j} x_i x_j \\ &\text{and in general} \\ \psi_r &= \left(\sum_{i_1 \leq i_2 \leq \dots \leq i_r} x_{i_1} x_{i_2} \dots x_{i_r} \right). \end{aligned} \tag{25}$$

These ψ_r -weights can also be computed as follows. The model can be written as

$$r_t = (\Phi(L))^{-1} e_t \text{ or}$$

$$r_t = \Psi(L) e_t$$

which implies that

$$(\Phi(L))^{-1} = \Psi(L) \text{ or}$$

$$\Phi(L)\Psi(L) = 1$$

that is

$$(1 - \phi_1 L - \phi_2 L^2 - \dots - \phi_p L^p)(1 + \psi_1 L + \psi_2 L^2 + \dots) = 1,$$

$$(1 + (\psi_1 - \phi_1)L + (\psi_2 - \psi_1\phi_1 - \phi_2)L^2 + \dots) = 1.$$

Thus comparing powers of L we get

$$L^1 : \psi_1 - \phi_1 = 0 \Rightarrow \psi_1 = \phi_1$$

$$L^2 : \psi_2 - \psi_1\phi_1 - \phi_2 = 0 \Rightarrow \psi_2 = \psi_1\phi_1 + \phi_2 = \phi_1^2 + \phi_2$$

and so on. In general we get for $j \geq 2$

$$L^j : \psi_j - \psi_{j-1}\phi_1 - \dots - \phi_j = 0 \Rightarrow \psi_j = \psi_{j-1}\phi_1 + \psi_{j-2}\phi_2 + \dots + \psi_{j-p}\phi_p.$$

(2) Wold has proved that any stationary process can be expressed in the form

$$\sum_{r=0}^{\infty} \psi_r e_{t-r}.$$

Thus in this sense it follows that the condition for stationarity for an $AR(p)$ process is that $|x_i| < 1$ for all the roots of $\Phi(\frac{1}{x}) = 0$. *This is also the condition of invertibility of an $AR(p)$ process.*

4.3.1 Identifying $AR(p)$ Models in Practice.

In applications the order p of an $AR(p)$ process is not known. It must be specified empirically. This is referred as the *order determination* of an $AR(p)$ process and has been

extensively studied in the time series literature. Two general approaches are available. The first approach is to use the partial autocorrelation function (PACF) and the second approach uses the information criterion function. In this study we focus on the first approach, via observation of the PACF and also the Autocorrelation Function (ACF). Now if the autocorrelation does not die off rapidly (i.e. it tapers off) and the partial autocorrelations, $\psi_{kk} = 0$ for $k > p$ then the process is identified as an $AR(p)$ process.

4.3.2 The $AR(p)$ Process and Regression Analysis.

Assuming that our series (r_1, \dots, r_T) is indeed generated by an AR process, we need to find its order p and the values of the parameters ϕ_1, \dots, ϕ_p to describe the process. If we know the order p then we can use regression analysis to estimate the parameters ϕ_1, \dots, ϕ_p since the equation

$$r_t = \phi_1 + \phi_1 r_{t-1} + \dots + \phi_p r_{t-p} + e_t$$

has precisely the form of a linear statistical regression model. Note, however the regressors (explanatory variables) $(r_{t-1}, \dots, r_{t-p})$ are stochastic variables. If the e_t are white noise, or even Gaussian white noise, then individual e_t represents random shock, which is added to the process at time t , and is independent of random variables at previous time points. Hence the regressors in a particular equation are independent of the error term. Thus we may estimate

$$\phi_p = \begin{pmatrix} \phi_0 \\ \phi_1 \\ \phi_2 \\ \vdots \\ \phi_p \end{pmatrix}$$

by the least squares (LS) method. Replacing the random variables by there observed values we get

$$\begin{aligned} r_{p+1} &= \phi_0 + \phi_1 r_p + \dots + \phi_p r_1 + e_{p+1} \\ r_{p+2} &= \phi_0 + \phi_1 r_{p+1} + \dots + \phi_p r_2 + e_{p+2} \\ &\vdots \\ r_T &= \phi_0 + \phi_1 r_{T-1} + \dots + \phi_p r_{T-p} + e_T \end{aligned}$$

or in matrix notation

$$\mathbf{y}_p = \mathbf{X}_p \boldsymbol{\phi}_p + \mathbf{e} \quad (26)$$

where

$$\mathbf{r}_p = \begin{pmatrix} r_{p+1} \\ r_{p+2} \\ \vdots \\ r_T \end{pmatrix}, \quad \mathbf{e} = \begin{pmatrix} e_{p+1} \\ e_{p+2} \\ \vdots \\ e_T \end{pmatrix} \quad (27)$$

and

$$\mathbf{X}_p = \begin{pmatrix} 1 & r_p & r_{p-1} & \cdots & r_1 \\ 1 & r_{p+1} & r_p & \cdots & r_2 \\ \vdots & \vdots & \vdots & \ddots & \vdots \\ 1 & r_{T-1} & r_{T-2} & \cdots & r_{T-p} \end{pmatrix}.$$

If the e_t and thus the r_t are normally distributed, the LS estimator of $\boldsymbol{\phi}$, with estimate

$$\hat{\boldsymbol{\phi}}_p = (\mathbf{X}_p' \mathbf{X}_p)^{-1} \mathbf{X}_p' \mathbf{r}_p$$

is consistent and asymptotically normally distributed. That is,

$$\sqrt{T}(\hat{\boldsymbol{\phi}}_p - \boldsymbol{\phi}_p) \rightarrow^d N(\mathbf{0}, \boldsymbol{\Sigma})$$

with the variance - covariance matrix

$$\boldsymbol{\Sigma} = \begin{pmatrix} \sigma^2/T & 0 \\ 0 & \boldsymbol{\Sigma}_{\hat{\boldsymbol{\phi}}_p} \end{pmatrix}$$

and

$$\Sigma_{\hat{\phi}_p} = \sigma^2 \begin{bmatrix} \gamma_0 & \gamma_1 & \gamma_2 & \cdots & \gamma_{p-1} \\ \gamma_1 & \gamma_0 & \gamma_1 & \cdots & \gamma_{p-2} \\ \vdots & \vdots & \vdots & \ddots & \vdots \\ \gamma_{p-1} & \gamma_{p-2} & \gamma_{p-3} & \cdots & \gamma_0 \end{bmatrix}.$$

The variance - covariance matrix Σ can be consistently estimated by

$$\hat{\Sigma} = \hat{\sigma}^2 (\mathbf{X}_p' \mathbf{X}_p)^{-1}$$

where

$$\hat{\sigma}^2 = \frac{(\mathbf{r}_p - \mathbf{X}_p \hat{\phi}_p)' (\mathbf{r}_p - \mathbf{X}_p \hat{\phi}_p)}{T - (2p + 1)},$$

is the LS estimate of the variance σ^2 of the white noise process e_t . Note that we use $T - (2p + 1)$ rather than $T - (p + 1)$ since there are only $T - p$ complete observations with which to estimate the $(p + 1)$ parameters in ϕ . Note that ϕ_0 is not the mean of the process r_t . Rather

$$\mu = \frac{\phi_0}{1 - \phi_1 - \cdots - \phi_p}.$$

Thus two possible estimates of the process mean μ are

$$\bar{r} = \frac{1}{T} \sum_{t=1}^T r_t$$

and

$$\hat{\mu} = \frac{\hat{\phi}_0}{1 - \hat{\phi}_1 - \cdots - \hat{\phi}_p}$$

where $\hat{\phi}_0, \hat{\phi}_1, \dots, \hat{\phi}_p$ are the least squares estimates. Under general conditions (e.g. the r_t are normally distributed) the corresponding estimates are consistent and have identical asymptotic normal distributions.. For instance

$$\sqrt{T}(\hat{\mu} - \mu) \rightarrow^d N[0, \sigma^2(1 - \phi_1 - \cdots - \phi_p)^{-2}].$$

Furthermore the estimators \bar{r} and $\hat{\mu}$ are asymptotically independent of the estimators $\hat{\phi}_1, \dots, \hat{\phi}_p$. This property is one reason why the estimation of ϕ_p is often discussed in terms of zero mean processes. In practice the sample mean may be subtracted from the data prior to estimating the other parameters. In this case the asymptotic distribution theory is not affected.

4.3.3 Estimation.

For a specified $AR(p)$ model the conditional least squares method starting with the $(p+1)th$ observation as discussed above is often used to estimate the parameters. Let

$$r_t = \phi_0 + \phi_1 r_{t-1} + \dots + \phi_p r_{t-p} + e_t, t = p+1, \dots, T \quad (28)$$

then the estimates are obtained by minimizing

$$\sum_{t=p+1}^T e_t^2$$

giving the least squares estimates $(\hat{\phi}_0, \hat{\phi}_1, \dots, \hat{\phi}_p)$ (see also (26) and (27)). The fitted model is

$$\hat{r}_t = \hat{\phi}_0 + \hat{\phi}_1 r_{t-1} + \dots + \hat{\phi}_p r_{t-p}, \quad (29)$$

and the estimated residual

$$\hat{e}_t = r_t - \hat{r}_t.$$

The series $\{\hat{e}_t\}_{t=p+1}^T$ is called the residual series from which we obtain

$$\hat{\sigma}^2 = \frac{1}{T-2p-1} \sum_{t=p+1}^T \hat{e}_t^2. \quad (30)$$

4.3.4 Model Checking.

A fitted model must be examined carefully to check for possible model inadequacy. If the model is adequate then the residual series should behave as a white noise. The *ACF* and the Ljung-Box statistics of the residuals can be used to check the closeness of $\{\hat{e}_t\}$ to a white noise. The Akaike *AIC* and Schwarz *SC* must also be examined.

For an $AR(p)$ process the Ljung-Box statistic $Q(m)$ follows asymptotically a chi-squared distribution with $m-p$ degrees of freedom. Here the number of degrees of freedom is modified to signify that, p , *AR* coefficients have been estimated. (Troskie (2002))

4.4 Empirical Study: Single Index

4.4.1 Introduction

Data in economics are frequently collected in form of time series. The intrinsic nature of a time series is that its observations are ordered in time and the modelling strategies of time series must take into account this property. This does not occur with cross-section data where the sequence of data points does not matter. Due to this order in time, it is likely that the value of a variable y at moment t reflects the past history of the series, that is, the observations of a time series are likely to be correlated. Since the observations are measurements of the same variable, it is usually said that y is correlated with itself, that is, it is autocorrelated.

Time Series Analysis is the set of statistical methodologies that analyze this kind of data. The main tool in Time Series Analysis is a model that should reproduce the past behavior of the series, exploiting its autocorrelation structure. The objectives of Time Series Analysis are basically to describe the regularity patterns present in the data and to forecast future observations. Since a pure time series model does not include explanatory variables, these forecasts of future observations are simply extrapolations of the observed series at the end of the sample. If we consider a single variable in our study, we shall construct what is called a univariate time series model. But if two or more variables are available, the possibility of dynamic interactions among them may be important. We can think, for instance, in economic variables such as consumption, investment and income that influence each other.

In this case, multivariate time series models can be constructed to take into account these relations among variables (Lutkepohl, 1991).

4.4.2 The Data

For this empirical analysis, the data set used is the SIDS batch.

4.4.3 Study Objectives

As mentioned in the introduction earlier, one of the objectives of time series analysis is to describe the regularity patterns present in the data. The objective in mind for this empirical study relates to this whereby we attempt to formulate single index time series models for our basket of 9 stocks. Previously, in the initial empirical study we had established regression models for the 9 stocks: $R_1, R_2, R_3, R_4, R_5, R_6, R_7, R_8$ and R_9 . Having observed the Durbin Watson (DW) statistic for all nine response variables, it was apparent that for most of them some form of autocorrelation was present. Hence the objective for this study is to improve on the found regression models by introducing time series components, if any found, to the existing nine regression models. The Box and Jenkins methodology will be adopted with some adjustments made with regards to time series modelling. A diagrammatic representation of this process is presented in the "Model Building Process" section of the previous chapter of "Multivariate Regression and Model Selection", the reader is referred to this to get an idea of the overall process.

The results obtained from this section will then be used as inputs in a later empirical study that constructs efficient frontiers of the single index time series models to be built.

4.4.4 Methodology

Box and Jenkins methodology is supported by Wold's Theorem, which states that any stationary stochastic process can be decomposed into a linearly deterministic part, $E(y_t|y_{t-1}, \dots, y_t)$ and a linearly indeterministic part, $\sum_{i=0}^{\infty} \pi_i \epsilon_{t-i}$. Choosing the correct ARMA process using the Box-Jenkins approach is as much art as science. The Box-Jenkins methodology for fitting an ARMA model to a time series consists of four steps:

1. Decide on the order of differencing that is needed to produce a stationary series $\{y_t\}$, i.e. Check for stationarity, if not, induce.
2. By inspecting the sample autocorrelations and partial autocorrelations, determine tentative values for p and q , i.e. Identify the AR and MA orders from the sample ACF and sample PACF by comparing them to the theoretical ACF and PACF.
3. Estimate the lag coefficients by maximum likelihood.
4. Using various diagnostics, check if the tentative model was indeed appropriate i.e. diagnostic testing. The adequacy of the model can be tested by examining the residuals or by over-fitting the obtained model.
5. Forecast (to be studied and explored another time, not to be investigated in this thesis)

In addition to this we will select a set of models based on prior considerations of maximum possible settings of p and q , estimate each possible model which minimises and maximises the chosen selection criteria based on goodness of fit considerations. Ultimately this should result in nine improved models that explain a larger percentage of variation in the dependent variables opposed to the percentage explained before by the ordinary

regression models. Again the principal criteria involved in deciding on the significance of these models are the R^2 , *Adjusted R^2* , *AIC*, *BIC*, the F-statistic and it's corresponding Probability/P-value. The following sub-section discloses the results realised from adopting the Box and Jenkins methodology and the application of the selection criteria. The time series package (TSP) used is EVIEWS 5 to implement the above outlined methodology.

4.4.5 Primary Findings

From the previous empirical study regression models were established for the nine financial stocks in our portfolio. The above outlined methodology will be applied for each of the regression models acquired from the previous chapter. We highlight the results from the above procedure for only one stock, namely Anglos. The results for the remaining eight stocks can be referred to at the end of this study in the List of Appendices section.

The first property to establish is that of stationarity. This is can clearly be established from viewing the residuals from the regression model or by implementing a simple Augmented Dickey Fuller test. We opt for the first option. Figure 4.4 below graphs the residuals.

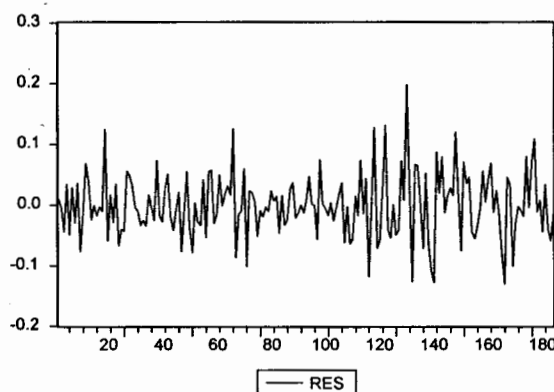


Figure 4.4: Graph of Residuals for Anglos

It is clear that the series is stationary about the mean. Towards the end of the series there is a clear indication of increased volatility, hence a presence of heteroskedasticity. Having established the stationarity property, next an examination of the autocorrelations (ACF) and partial autocorrelations (PACF) is required in order to establish whether any autoregressive (AR) or moving average (MA) components are required to be fit. Figure 4.5 below presents the correlogram constructed in Eviews5 of the residuals from the regression model. Exploring the correlogram below and paying attention to the spikes at the various lags for the ACF and PACF, it is clear that at lags 2 and 6 the spikes are significant at those points for both the PACF and ACF.

ACF	PACF		AC	PAC	Q-Stat	Prob
		1	0.054	0.054	0.2982	0.585
		2	-0.165	-0.168	3.1555	0.206
		3	0.041	0.063	3.3373	0.342
		4	0.094	0.061	4.2856	0.369
		5	-0.044	-0.039	4.4926	0.481
		6	-0.174	-0.151	7.8095	0.252
		7	0.032	0.036	7.9249	0.339
		8	0.038	-0.019	8.0860	0.425
		9	-0.079	-0.053	8.7894	0.457
		10	-0.075	-0.049	9.4330	0.492

Figure 4.5: Correlogram for Anglos

At first one would assume from the theory presented above that we would fit the regression model with components: AR(2), AR(6), MA(2) and MA(6). This is not the case if one of the objectives is to maintain simplicity in model structure. The solution is quite simple, as mentioned before in "Methodology", we will select a set of models based on prior considerations of maximum possible settings of p and q , estimate each possible model which minimises or maximises the chosen selection criteria based on goodness of fit considerations.

The reader is also referred to the correlograms constructed from the residuals from the regressions of the remaining eight stocks in the List of Appendices section under Appendix 2A. Having observed all the correlograms, the following components are fitted:

We fit an AR(2) component to regression model 1. The correlogram of the residuals from Regression model 2 illustrate random scatter, thus no time series component is required. We fit an AR(1) component to both regression models 3 and 4, since both are very significant at lag 1 for the 5% significance level. Regression models 5, 6 and 7 exhibit random scatter in the illustration of their residuals. In the correlogram for the residuals of regression model 8, the Q-statistic becomes really significant at lag 6, thus we fit an AR(6) component. Finally for model 9 we fit an AR(2) component. Table 4.1 below illustrates a summary of the model selection criteria from fitting the new models

Time Series Models	Model Selection Criteria						
	Adj-R ²	S.E. of reg	D-W stat	AIC	BIC	F-stat	Prob(F-stat)
R1 = const+ β_1 'JSE+ ϕ_2 'R1 _{t-2} + e _{R1}	0.6877	0.0558	2.0342	-2.9187	-2.8687	216.7925	0.0000
R2 = const+ β_2 'JSE+e ₂	0.1955	0.1230	2.1197	-1.3440	-1.3109	49.1156	0.0000
R3= const+ β_3 'JSE+ ϕ_1 'R3 _{t-1} + e _{R3}	0.2182	0.0903	1.9724	-1.9555	-1.9057	28.4901	0.0000
R4= const+ β_4 'JSE+ ϕ_1 'R4 _{t-1} + e _{R4}	0.4133	0.0609	1.9546	-2.7434	-2.6936	70.3988	0.0000
R5= const+ β_5 'JSE+e ₅	0.0434	0.0820	1.9712	-2.1549	-2.1218	9.9893	0.0018
R6= const+ β_6 'JSE+e ₆	0.3276	0.0908	1.8826	-1.9503	-1.9172	97.4576	0.0000
R7 = const+ β_7 'JSE+e ₇	0.3861	0.0785	1.8729	-2.2403	-2.2072	115.3412	0.0000
R8= const+ β_8 'JSE+ ϕ_1 'R8 _{t-1} + e _{R8}	0.2940	0.0653	2.1354	-2.6030	-2.5522	40.9743	0.0000
R9= const+ β_9 'JSE+ ϕ_2 'R9 _{t-2} + e _{R9}	0.2336	0.0804	2.0377	-2.1890	-2.1390	30.8703	0.0000

Table 4.1: Summary Stats for time series Models

Examining the above table it is clear that once again all the models are highly significant. The Adj-R² statistic has increased for stocks R₁, R₃, R₄, R₈ and R₉. This is expected due to the addition of an extra significant explanatory variable to each of the models, hence we are explaining more variation than before. For these same 5 models, a vast improvement in the D-W statistic is observed from before, thus is a probable indication of the extreme reduction in autocorrelation in the models. The other 4 models have remained the same from

before due to the absence of significant autocorrelation in the residuals of their regression models. In the new models for R_1, R_3, R_4, R_8 and R_9 , improvements in all five's SE of regression is also noted. The AIC and BIC's have remained relatively the same. Next we examine the significance of the new updated coefficients, presented in table 4.2 below.

Time Series Models	Coefficient of JSE (β_i)			
	Coefficient (β_i)	Std. Error	t-Statistic	Prob.
$R1 = \text{const} \cdot \beta_1 \cdot \text{JSE} + \phi_1 \cdot R1_{1,z} + e_{R1}$	1.4200	0.0679	20.9161	0.0000
$R2 = \text{const} \cdot \beta_2 \cdot \text{JSE} + e_{R2}$	1.0359	0.1478	7.0083	0.0000
$R3 = \text{const} \cdot \beta_3 \cdot \text{JSE} + \phi_1 \cdot R3_{1,z} + e_{R3}$	0.7029	0.1047	6.7127	0.0000
$R4 = \text{const} \cdot \beta_4 \cdot \text{JSE} + \phi_1 \cdot R4_{1,z} + e_{R4}$	0.8262	0.0727	11.3654	0.0000
$R5 = \text{const} \cdot \beta_5 \cdot \text{JSE} + e_{R5}$	0.3115	0.0985	3.1606	0.0018
$R6 = \text{const} \cdot \beta_6 \cdot \text{JSE} + e_{R6}$	1.0776	0.1092	9.8721	0.0000
$R7 = \text{const} \cdot \beta_7 \cdot \text{JSE} + e_{R7}$	1.0141	0.0944	10.7397	0.0000
$R8 = \text{const} \cdot \beta_8 \cdot \text{JSE} + \phi_1 \cdot R8_{1,z} + e_{R8}$	0.7327	0.0773	9.4758	0.0000
$R9 = \text{const} \cdot \beta_9 \cdot \text{JSE} + \phi_2 \cdot R9_{1,z} + e_{R9}$	0.7450	0.0965	7.7202	0.0000
Time Series Models	Coefficient of AR(p) (ϕ_i)			
	Coefficient (β_i)	Std. Error	t-Statistic	Prob.
$R1 = \text{const} \cdot \beta_1 \cdot \text{JSE} + \phi_1 \cdot R1_{1,z} + e_{R1}$	-0.1386	0.0722	-1.9200	0.0563
$R2 = \text{const} \cdot \beta_2 \cdot \text{JSE} + e_{R2}$	0.0000	0.0000	0.0000	0.0000
$R3 = \text{const} \cdot \beta_3 \cdot \text{JSE} + \phi_1 \cdot R3_{1,z} + e_{R3}$	-0.2923	0.0686	-4.2597	0.0000
$R4 = \text{const} \cdot \beta_4 \cdot \text{JSE} + \phi_1 \cdot R4_{1,z} + e_{R4}$	-0.1516	0.0712	-2.1297	0.0345
$R5 = \text{const} \cdot \beta_5 \cdot \text{JSE} + e_{R5}$	0.0000	0.0000	0.0000	0.0000
$R6 = \text{const} \cdot \beta_6 \cdot \text{JSE} + e_{R6}$	0.0000	0.0000	0.0000	0.0000
$R7 = \text{const} \cdot \beta_7 \cdot \text{JSE} + e_{R7}$	0.0000	0.0000	0.0000	0.0000
$R8 = \text{const} \cdot \beta_8 \cdot \text{JSE} + \phi_1 \cdot R8_{1,z} + e_{R8}$	-0.2374	0.0708	-3.3511	0.0010
$R9 = \text{const} \cdot \beta_9 \cdot \text{JSE} + \phi_2 \cdot R9_{1,z} + e_{R9}$	-0.1862	0.0715	-2.6047	0.0099

Table 4.2: Summary of Coefficients β and ϕ

All the above coefficients are significant at the 5% level of significance. Since the models for R_2, R_5, R_6 and R_7 have remain unchanged, we only have to confirm the absence of autocorrelation from the other 5 new models. This is achieved by examining the residuals of the new time series models of these variables. The correlogram for Anglos, R_1 , is illustrated below in figure 4.6. The remainder of the four correlograms can be found in Appendix 2B.

ACF	PACF		AC	PAC	Q-Stat	Prob
		1	0.072	0.072	0.5362	0.464
		2	-0.018	-0.024	0.5713	0.752
		3	0.041	0.044	0.7469	0.862
		4	0.050	0.044	1.0199	0.907
		5	-0.037	-0.042	1.1662	0.948
		6	-0.163	-0.159	4.0739	0.667
		7	0.011	0.029	4.0870	0.770
		8	0.013	0.006	4.1063	0.847
		9	-0.078	-0.064	4.7880	0.852
		10	-0.062	-0.041	5.2215	0.876

Figure 4.6: Residuals for Anglos

Scrutinizing the above correlogram and the remaining four in appendix 2B, it is clear that all five portray a situation of random scatter, that the Q-statistic is insignificant at all lags. This implies that we accept the null $H_0 : \rho = 0$, which is interpreted as no autocorrelation present. Therefore we can terminate our model building phase here.





















4.4.6 Conclusions

Having completed the Box and Jenkins methodology, the following is concluded: that significant time series models have been found for five out of the nine stocks of our portfolio. Furthermore the time series models acquired appear to provide better explanatory power in terms of higher R^2 values and lower error variances of the regression than the regression models for stocks R_1 , R_3 , R_4 , R_8 and R_9 . Also the beta estimates acquired seem to be more stable for the five stocks as well. The betas and regression SEs computed for the nine stocks will be used as inputs for the computation of the efficient frontier for the single index time series model. The variance of JSE and the mean log returns of the nine stocks will also be required as inputs for the computation of the regression method's mean-variance frontier.

The next empirical study conducts a similar investigation, but in a multiple index setting.

4.5 List of Appendices





















4.5.1 Appendix 2A: Correlograms for the residuals from the regression models for the remaining eight stocks

ACF	PACF	AC	PAC	Q-Stat	Prob
		1 -0.049	-0.049	0.2533	0.615
		2 0.041	0.039	0.4342	0.805
		3 0.171	0.175	3.5313	0.317
		4 -0.066	-0.051	3.9928	0.407
		5 0.021	-0.000	4.0410	0.544
		6 -0.032	-0.058	4.1556	0.656
		7 -0.076	-0.064	4.7990	0.684
		8 0.005	-0.004	4.8021	0.779
		9 -0.012	0.012	4.8181	0.850
		10 0.035	0.056	4.9551	0.894

Correlogram of JD Group

ACF	PACF	AC	PAC	Q-Stat	Prob	
		1	0.074	0.074	0.5710	0.450
		2	0.006	0.001	0.5752	0.750
		3	-0.077	-0.078	1.2098	0.751
		4	-0.100	-0.090	2.2857	0.683
		5	0.127	0.144	4.0376	0.544
		6	0.122	0.102	5.6690	0.461
		7	0.086	0.054	6.4798	0.485
		8	0.014	0.010	6.5014	0.591
		9	-0.136	-0.104	8.5995	0.475
		10	-0.078	-0.053	9.2869	0.505

Correlogram of SA-Eagle

ACF	PACF	AC	PAC	Q-Stat	Prob	
		1	-0.079	-0.079	0.6572	0.418
		2	0.110	0.104	1.9176	0.383
		3	-0.044	-0.029	2.1272	0.546
		4	-0.172	-0.192	5.3004	0.258
		5	-0.170	-0.199	8.4323	0.134
		6	-0.172	-0.183	11.683	0.069
		7	0.058	0.049	12.057	0.099
		8	-0.105	-0.117	13.294	0.102
		9	-0.008	-0.146	13.302	0.149
		10	0.137	0.046	15.459	0.116

Correlogram of TigBrands

ACF	PACF	AC	PAC	Q-Stat	Prob	
		1	-0.319	-0.319	10.582	0.001
		2	0.006	-0.107	10.586	0.005
		3	0.031	-0.001	10.689	0.014
		4	0.059	0.080	11.066	0.026
		5	-0.188	-0.158	14.903	0.011
		6	0.059	-0.058	15.286	0.018
		7	-0.080	-0.112	15.996	0.025
		8	0.077	0.033	16.661	0.034
		9	-0.119	-0.086	18.252	0.032
		10	0.093	0.010	19.250	0.037





















Correlogram for Pick & Pay

ACF	PACF	AC	PAC	Q-Stat	Prob	
		1	0.059	0.059	0.3624	0.547
		2	-0.017	-0.020	0.3926	0.822
		3	0.110	0.112	1.6666	0.644
		4	0.063	0.050	2.0884	0.720
		5	-0.043	-0.046	2.2925	0.807
		6	-0.033	-0.038	2.4137	0.878
		7	-0.093	-0.105	3.3718	0.849
		8	0.061	0.079	3.7817	0.876
		9	0.066	0.070	4.2709	0.893
		10	-0.057	-0.039	4.6369	0.914

Correlogram of Sappi

ACF	PACF	AC	PAC	Q-Stat	Prob	
		1	0.002	0.002	0.980	
		2	-0.328	-0.328	11.302	0.004
		3	0.049	0.057	11.559	0.009
		4	0.118	0.010	13.043	0.011
		5	-0.113	-0.091	14.432	0.013
		6	-0.035	0.009	14.568	0.024
		7	0.152	0.094	17.128	0.017
		8	0.069	0.070	17.665	0.024
		9	-0.114	-0.033	19.131	0.024
		10	-0.027	0.001	19.214	0.038

Correlogram of Tongaat

ACF	PACF	AC	PAC	Q-Stat	Prob	
		1	-0.244	-0.244	6.1825	0.013
		2	0.056	-0.003	6.5174	0.038
		3	-0.116	-0.110	7.9489	0.047
		4	0.040	-0.014	8.1228	0.087
		5	-0.019	-0.010	8.1617	0.148
		6	0.032	0.015	8.2725	0.219
		7	-0.135	-0.131	10.292	0.173
		8	-0.111	-0.193	11.678	0.166
		9	0.128	0.065	13.522	0.140
		10	0.016	0.043	13.552	0.194

Correlogram of Remgro

ACF		PACF		AC	PAC	Q-Stat	Prob	
				1	-0.006	-0.006	0.0036	0.952
				2	-0.041	-0.041	0.1814	0.913
				3	-0.096	-0.096	1.1554	0.764
				4	0.042	0.039	1.3412	0.854
				5	-0.034	-0.042	1.4670	0.917
				6	-0.029	-0.036	1.5615	0.955
				7	-0.030	-0.027	1.6643	0.976
				8	0.009	-0.003	1.6740	0.989
				9	-0.009	-0.016	1.6841	0.996
				10	-0.019	-0.024	1.7266	0.998

Correlogram of Sasol

4.5.2 Appendix 2B: Correlograms for the residuals from the Time Series models for R_3 , R_4 , R_8 and R_9 stocks

Correlograms:

ACF	PACF	AC	PAC	Q-Stat	Prob	
		1	-0.067	-0.067	0.4617	0.497
		2	-0.084	-0.089	1.2049	0.547
		3	0.064	0.052	1.6369	0.651
		4	0.023	0.024	1.6947	0.792
		5	-0.186	-0.176	5.4534	0.363
		6	-0.013	-0.039	5.4727	0.485
		7	-0.049	-0.086	5.7338	0.571
		8	0.031	0.037	5.8380	0.665
		9	-0.082	-0.083	6.5928	0.679
		10	0.067	0.038	7.1070	0.715

Residuals for Pick and Pay

ACF	PACF	AC	PAC	Q-Stat	Prob	
		1	-0.006	-0.006	0.0036	0.952
		2	-0.041	-0.041	0.1814	0.913
		3	-0.096	-0.096	1.1554	0.764
		4	0.042	0.039	1.3412	0.854
		5	-0.034	-0.042	1.4670	0.917
		6	-0.029	-0.036	1.5615	0.955
		7	-0.030	-0.027	1.6643	0.976
		8	0.009	-0.003	1.6740	0.989
		9	-0.009	-0.016	1.6841	0.996
		10	-0.019	-0.024	1.7266	0.998

Residuals for Tigerbrands

ACF	PACF		AC	PAC	Q-Stat	Prob
		1	-0.107	-0.107	1.1867	0.276
		2	0.007	-0.005	1.1914	0.551
		3	-0.110	-0.111	2.4682	0.481
		4	0.022	-0.002	2.5201	0.641
		5	-0.011	-0.010	2.5332	0.771
		6	0.028	0.015	2.6210	0.855
		7	-0.161	-0.158	5.4794	0.602
		8	-0.117	-0.160	7.0041	0.536
		9	0.120	0.096	8.6410	0.471
		10	0.018	0.006	8.6767	0.563

Residuals for Remgro

ACF	PACF	AC	PAC	Q-Stat	Prob	
		1	0.011	0.011	0.912	
		2	-0.147	-0.148	2.2946	0.317
		3	0.032	0.036	2.4017	0.493
		4	0.063	0.041	2.8312	0.586
		5	-0.091	-0.085	3.7305	0.589
		6	0.001	0.018	3.7305	0.713
		7	0.127	0.101	5.5015	0.599
		8	0.063	0.066	5.9463	0.653
		9	-0.078	-0.042	6.6282	0.676
		10	-0.027	-0.023	6.7108	0.752

Residuals for Tongaat

4.6 Empirical Study: Multiple Index

4.6.1 The Data

For this empirical analysis, the data used is the MIDS batch.

4.6.2 Study Objectives

Similar to the objective in the preceding empirical investigation, once again we aim to describe the regularity patterns present in the data. As shown previously, we had established time series models for 5 of the 9 response variables in the single index environ. The objective for this study is to improve on the multiple index regression models by introducing time series components, if any exist, to the currently 9 multiple regression models. Once again the Box and Jenkins methodology will be adopted with some adjustments made with regards to time series modelling. A diagrammatic representation of this process is presented in the "Model Building Process" section of the previous chapter of "Multivariate Regression and Model Selection", the reader is again referred to this to get an idea of the overall process.

4.6.3 Methodology

The methodology adopted in the single index scenario will also be adopted for the multiple index setting. Recall that the Box-Jenkins methodology for fitting an ARMA model to a time series consists of four steps:

Exploring all 9 correlograms, it turns out once again like in the single index case that response variables R_1, R_3, R_4, R_8 and R_9 all have time series components. What is even more coincidental is that they all exhibit the same time series components as in the single index setting. In other words, the Q-statistic is significant at lags that were found significant in the single index environ. Hence for these 5 MIR models we fit the same respective time series components which were fit before for the single index models, which are:

- an AR(2) component for R_1
- an AR(1) component for R_3
- an AR(1) component for R_4
- an AR(6) component for R_8
- and an AR(2) component for R_9 .

The remaining 4 responses remain in their MIR model form, due to their correlograms expressing random scatter, such that the Q-statistic is insignificant at all lags. This means that we accept the null hypothesis of no autocorrelation, hence no time series components for these 4 responses. We now refit new models for R_1, R_3, R_4, R_8 and R_9 with their respective time series components and observe the updated model selection criterion, which is illustrated in table 4.4 below:

Multiple Time Series Models	Model Selection Criteria						
	Adj-R ²	S.E. of reg	D-W stat	AIC	BIC	F-stat	Prob(F-stat)
Model for R1 (Model 1)	0.7112	0.0537	2.0385	-2.9871	-2.9038	121.6846	0.0000
Model for R2 (Model 2)	0.2111	0.1218	2.1906	-1.3537	-1.2875	18.6621	0.0000
Model for R3 (Model 3)	0.3027	0.0853	1.9841	-2.0600	-1.9770	22.3765	0.0000
Model for R4 (Model 4)	0.4168	0.0607	1.9555	-2.7395	-2.6564	36.2008	0.0000
Model for R5 (Model 5)	0.0554	0.0815	2.0528	-2.1577	-2.0915	4.8720	0.0027
Model for R6 (Model 6)	0.3774	0.0874	1.9198	-2.0174	-1.9512	41.0095	0.0000
Model for R7 (Model 7)	0.3798	0.0777	1.8304	-2.2523	-2.1861	41.4142	0.0000
Model for R8 (Model 8)	0.3364	0.0633	2.0863	-2.6548	-2.5702	25.3316	0.0000
Model for R9 (Model 9)	0.2565	0.0792	2.0565	-2.2094	-2.1260	17.9048	0.0000

Table 4.4: Summary Stats for MIR time series models

Before any further analysis of the above results, table 4.5 below presents the model representations adopted for each of the response variables R_1 through to R_9 .

Multiple Reg Models	Definition Of Multiple Reg Structure
Model for R1 (Model 1)	$R_1 = \text{const} + \beta_1 \text{'Implats'} + \beta_2 \text{'JSE'} + \beta_3 \text{'Palam'} + \phi_2 R_{1,t-2} + e_1$
Model for R2 (Model 2)	$R_2 = \text{const} + \beta_1 \text{'Implats'} + \beta_2 \text{'JSE'} + \beta_3 \text{'Palam'} + e_2$
Model for R3 (Model 3)	$R_3 = \text{const} + \beta_1 \text{'Implats'} + \beta_2 \text{'JSE'} + \beta_3 \text{'Palam'} + \phi_1 R_{3,t-1} + e_3$
Model for R4 (Model 4)	$R_4 = \text{const} + \beta_1 \text{'Implats'} + \beta_2 \text{'JSE'} + \beta_3 \text{'Palam'} + \phi_1 R_{4,t-1} + e_4$
Model for R5 (Model 5)	$R_5 = \text{const} + \beta_1 \text{'Implats'} + \beta_2 \text{'JSE'} + \beta_3 \text{'Palam'} + e_5$
Model for R6 (Model 6)	$R_6 = \text{const} + \beta_1 \text{'Implats'} + \beta_2 \text{'JSE'} + \beta_3 \text{'Palam'} + e_6$
Model for R7 (Model 7)	$R_7 = \text{const} + \beta_1 \text{'Implats'} + \beta_2 \text{'JSE'} + \beta_3 \text{'Palam'} + e_7$
Model for R8 (Model 8)	$R_8 = \text{const} + \beta_1 \text{'Implats'} + \beta_2 \text{'JSE'} + \beta_3 \text{'Palam'} + \phi_6 R_{8,t-6} + e_8$
Model for R9 (Model 9)	$R_9 = \text{const} + \beta_1 \text{'Implats'} + \beta_2 \text{'JSE'} + \beta_3 \text{'Palam'} + \phi_2 R_{9,t-2} + e_9$

Table 4.5: MIR Time Series model representations

Exploring table 4.4 above, it clear that once again all the models are highly significant. The $Adj-R^2$ statistic has increased for stocks R_1 , R_3 , R_4 , R_8 and R_9 , compared with their MIR models. This is expected due to the addition of an extra significant explanatory variable to each of the models, hence we are explaining more variation than before. With regards to the D-W statistic for these 5 responses, a very interesting scenario has emerged. It seems that the D-W stat has improved for response variables R_3 , R_4 and R_8 , implying less auto-

correlation, while worsening for the other two, implying even more autocorrelation. For the multiple index case, the AICs of all 5 responses have improved. The BIC like the D-W statistic, also produced similar results. The other 4 models, namely R_2 , R_5 , R_6 and R_7 have remained the same from before due to the absence of significant autocorrelation in the residuals of their regression models.

Next we confirm the significance of the new updated coefficients. The reader is referred to appendix 2D for the computational results for the beta coefficients of the 5 new time series models constructed for R_1 , R_3 , R_4 , R_8 and R_9 . Observing the 5 tables in appendix 2D, it is clear that all 5 time series component's betas are significant at the 5% level of significance. The beta coefficients of the explanatory variables remain adequate from before. Since the models for R_2 , R_5 , R_6 and R_7 have remain unchanged, we only have to confirm the absence of autocorrelation from the other 5 new models. This is achieved by examining the residuals of the new MIR time series models of these variables. The correlogram for Anglos, R_1 , is illustrated below in figure 4.7. The remainder of the four correlograms can be found in Appendix 2E.

	AC	PAC	Q-Stat	Prob
1	0.089	0.089	0.8273	0.363
2	-0.047	-0.055	1.0578	0.589
3	-0.058	-0.049	1.4181	0.701
4	0.027	0.035	1.4946	0.828
5	-0.076	-0.088	2.1156	0.833
6	-0.120	-0.107	3.6979	0.717
7	0.036	0.054	3.8376	0.798
8	-0.066	-0.098	4.3221	0.827
9	-0.101	-0.095	5.4750	0.791
10	-0.045	-0.029	5.7100	0.839

Fig 4.7: Residuals of R1

Examining figure 4.7 above and the remaining four correlograms in appendix 2e, it is clear

that all five portray a situation of random scatter, that the Q-statistic is insignificant at all lags. This implies that we accept the null $H_0 : \rho = 0$, which is interpreted as no autocorrelation present. Therefore we can terminate our Box - Jenkins model building phase here.

4.6.5 Conclusions

Having completed the Box and Jenkins methodology, the following conclusion is arrived at: that it can be concluded that significant MIR time series models have been found for 5 out of the 9 stocks of our portfolio. Furthermore the MIR time series models acquired appear to provide better explanatory power in terms of higher R^2 values and lower error variances of the regression than their regression counterparts for stocks R_1, R_3, R_4, R_8 and R_9 . In addition the beta estimates acquired seem to be more stable for the five responses. The betas and regression SEs computed for the nine stocks will be used as inputs for the computation of the efficient frontier for the multiple index time series model. The variance of JSE and the mean log returns of the nine stocks will also be required as inputs for the computation of the method's mean-variance frontier.

The next chapter looks at State Space Modelling, a fairly new form of modelling. The previously found regression and time series models are now to be re-fitted, but in State Space form. This will be explored in the Empirical Study sections of the subsequent chapter.

4.7 List of Appendices

4.7.1 Appendix 2C: Correlograms for the residuals from the 9 MIR time series models

Correlogram

	AC	PAC	Q-Stat	Prob
1	0.075	0.075	0.5874	0.443
2	-0.220	-0.227	5.6847	0.058
3	-0.045	-0.009	5.9042	0.116
4	0.082	0.039	6.6222	0.157
5	-0.071	-0.100	7.1718	0.208
6	-0.122	-0.087	8.8022	0.185
7	0.066	0.056	9.2907	0.232
8	-0.039	-0.113	9.4602	0.305
9	-0.101	-0.067	10.611	0.303
10	-0.050	-0.060	10.897	0.366

Correlogram for MIR-R1

	AC	PAC	Q-Stat	Prob
1	-0.245	-0.245	6.2327	0.013
2	0.066	0.006	6.6869	0.035
3	-0.132	-0.122	8.5330	0.036
4	0.071	0.012	9.0772	0.059
5	0.026	0.055	9.1507	0.103
6	0.038	0.046	9.3119	0.157
7	-0.143	-0.125	11.580	0.115
8	-0.099	-0.170	12.681	0.123
9	0.118	0.071	14.248	0.114
10	0.013	0.036	14.268	0.161

Correlogram for MIR-R4

	AC	PAC	Q-Stat	Prob
1	0.010	0.010	0.0104	0.919
2	-0.075	-0.075	0.6033	0.740
3	-0.109	-0.108	1.8590	0.602
4	0.003	-0.002	1.8597	0.762
5	-0.050	-0.067	2.1310	0.831
6	-0.035	-0.047	2.2635	0.894
7	-0.009	-0.018	2.2721	0.943
8	0.035	0.016	2.4107	0.966
9	-0.041	-0.054	2.6032	0.978
10	-0.008	-0.010	2.6100	0.989

Correlogram for MIR-R7

	AC	PAC	Q-Stat	Prob
1	-0.096	-0.096	0.9527	0.329
2	0.069	0.060	1.4540	0.483
3	0.171	0.185	4.5423	0.209
4	-0.121	-0.096	6.1169	0.191
5	0.061	0.016	6.5253	0.258
6	-0.019	-0.027	6.5657	0.363
7	-0.047	-0.021	6.8079	0.449
8	-0.036	-0.067	6.9560	0.541
9	-0.029	-0.018	7.0525	0.632
10	0.040	0.053	7.2316	0.703

Correlogram for MIR-R2

	AC	PAC	Q-Stat	Prob
1	0.036	0.036	0.1324	0.716
2	-0.012	-0.013	0.1475	0.929
3	-0.084	-0.083	0.8912	0.828
4	-0.122	-0.117	2.4930	0.646
5	0.116	0.124	3.9526	0.556
6	0.095	0.081	4.9430	0.551
7	0.068	0.046	5.4587	0.604
8	0.048	0.051	5.7140	0.679
9	-0.165	-0.132	8.7899	0.457
10	-0.076	-0.057	9.4544	0.490

Correlogram for MIR-R5

	AC	PAC	Q-Stat	Prob
1	-0.055	-0.055	0.3094	0.578
2	0.037	0.034	0.4506	0.798
3	-0.061	-0.058	0.8462	0.838
4	-0.149	-0.157	3.2137	0.523
5	-0.153	-0.171	5.7400	0.332
6	-0.185	-0.214	9.4826	0.148
7	0.100	0.056	10.587	0.158
8	-0.152	-0.195	13.157	0.107
9	-0.011	-0.144	13.170	0.155
10	0.142	0.049	15.468	0.116

Correlogram for MIR-R8

	AC	PAC	Q-Stat	Prob
1	-0.288	-0.288	8.6406	0.003
2	-0.047	-0.141	8.8688	0.012
3	0.093	0.041	9.7874	0.020
4	0.044	0.089	9.9904	0.041
5	-0.148	-0.105	12.350	0.030
6	0.102	0.031	13.495	0.036
7	-0.091	-0.089	14.412	0.044
8	-0.006	-0.036	14.416	0.072
9	-0.050	-0.080	14.699	0.100
10	0.041	-0.008	14.889	0.136

Correlogram for MIR-R3

	AC	PAC	Q-Stat	Prob
1	0.042	0.042	0.1794	0.672
2	-0.041	-0.043	0.3538	0.838
3	0.150	0.154	2.7291	0.435
4	-0.011	-0.028	2.7428	0.602
5	-0.102	-0.089	3.8751	0.568
6	-0.028	-0.044	3.9580	0.682
7	-0.105	-0.107	5.1670	0.640
8	0.018	0.056	5.2040	0.736
9	0.028	0.024	5.2898	0.808
10	-0.010	0.014	5.3008	0.870

Correlogram for MIR-R6

	AC	PAC	Q-Stat	Prob
1	0.010	0.010	0.0111	0.916
2	-0.325	-0.325	11.079	0.004
3	0.024	0.036	11.142	0.011
4	0.077	-0.033	11.781	0.019
5	-0.105	-0.097	12.975	0.024
6	0.001	0.023	12.976	0.043
7	0.178	0.126	16.477	0.021
8	0.032	0.043	16.589	0.035
9	-0.097	0.001	17.649	0.039
10	0.008	0.024	17.656	0.061

Correlogram for MIR-R9

4.7.2 Appendix 2D: The MIR time series Models: Summary of Beta coefficients

Variable	Coefficient	Std. Error	t-Statistic	Prob.
C	-0.002518	0.003337	-0.754742	0.4513
X5	0.136887	0.035997	3.802772	0.0002
X6	1.225833	0.080131	15.29779	0.0000
X7	0.054991	0.034606	1.589043	0.1137
AR(2)	-0.171341	0.072166	-2.374250	0.0186

MIR Time Series - R1

Variable	Coefficient	Std. Error	t-Statistic	Prob.
C	0.008191	0.003812	2.148700	0.0329
X5	-0.068638	0.041185	-1.666589	0.0972
X6	0.918577	0.089979	10.20876	0.0000
X7	-0.017588	0.039720	-0.442790	0.6584
AR(1)	-0.158388	0.072198	-2.193817	0.0294

MIR Time Series - R4

Variable	Coefficient	Std. Error	t-Statistic	Prob.
C	0.001149	0.004833	0.237756	0.8123
X5	-0.043950	0.052890	-0.830981	0.4070
X6	0.721689	0.117413	6.146598	0.0000
X7	0.139824	0.050738	2.755779	0.0064
AR(2)	-0.194171	0.072013	-2.696321	0.0076

MIR Time Series - R9

Variable	Coefficient	Std. Error	t-Statistic	Prob.
C	0.007288	0.004852	1.502010	0.1347
X5	-0.282798	0.056668	-4.990389	0.0000
X6	1.004090	0.123250	8.146755	0.0000
X7	0.057354	0.054256	1.057109	0.2918
AR(1)	-0.283265	0.069782	-4.059266	0.0001

MIR Time Series - R3

Variable	Coefficient	Std. Error	t-Statistic	Prob.
C	0.002977	0.003773	0.789075	0.4311
X5	-0.154592	0.042193	-3.663912	0.0003
X6	0.927717	0.091886	10.09635	0.0000
X7	-0.029512	0.041345	-0.713804	0.4762
AR(6)	-0.234603	0.071266	-3.291930	0.0012

MIR Time Series - R8

4.7.3 Appendix 2E: Correlograms of the Final MIR- time series Models for R_3 , R_4 , R_8 and R_9

	AC	PAC	Q-Stat	Prob
1	-0.057	-0.057	0.3339	0.563
2	-0.111	-0.114	1.6248	0.444
3	0.105	0.093	2.7974	0.424
4	0.037	0.037	2.9454	0.567
5	-0.136	-0.113	4.9403	0.423
6	0.042	0.027	5.1304	0.527
7	-0.075	-0.108	5.7613	0.568
8	-0.040	-0.021	5.9411	0.654
9	-0.052	-0.074	6.2453	0.715
10	0.034	0.021	6.3757	0.783

Residuals of R3

	AC	PAC	Q-Stat	Prob
1	-0.078	-0.078	0.6310	0.427
2	-0.023	-0.029	0.6858	0.710
3	-0.083	-0.088	1.4170	0.702
4	-0.115	-0.132	2.8415	0.585
5	-0.171	-0.204	5.9921	0.307
6	0.019	-0.040	6.0324	0.420
7	0.115	0.079	7.5043	0.378
8	-0.137	-0.180	9.5928	0.295
9	-0.076	-0.171	10.243	0.331
10	0.148	0.101	12.752	0.238

Residuals of R8

	AC	PAC	Q-Stat	Prob
1	-0.102	-0.102	1.0750	0.300
2	0.011	0.001	1.0877	0.581
3	-0.125	-0.125	2.7336	0.435
4	0.066	0.042	3.1962	0.526
5	0.044	0.056	3.4062	0.638
6	0.040	0.036	3.5776	0.734
7	-0.169	-0.153	6.7305	0.457
8	-0.107	-0.136	8.0145	0.432
9	0.111	0.097	9.4110	0.400
10	0.014	-0.004	9.4337	0.491

Residuals of R4

	AC	PAC	Q-Stat	Prob
1	-0.029	-0.029	0.1673	0.683
2	0.009	0.008	0.1840	0.912
3	-0.024	-0.024	0.3037	0.959
4	0.056	0.054	0.9310	0.920
5	-0.107	-0.104	3.2696	0.659
6	0.024	0.018	3.3910	0.758
7	0.048	0.054	3.8652	0.795
8	-0.008	-0.014	3.8783	0.868
9	0.014	0.025	3.9184	0.917
10	-0.039	-0.050	4.2419	0.936

Residuals of R9

Chapter 5

State Space Modelling and Empirical Studies

5.1 Introduction

Dynamic systems can be represented in a general form known as the state space form. Many time series models, including the classical linear regression model and ARIMA models, can be written and estimated as special cases of a state space specification. There are two main benefits to representing a dynamic system in state space form. First, the state space allows unobserved variables (known as the state variables) to be incorporated into, and estimated along with, the observable model. Second, state space models can be estimated using a powerful recursive algorithm known as the Kalman filter. The Kalman filter is used both to evaluate the likelihood function and to forecast and smooth the unobserved state variables.

State space models have been applied in the econometrics literature to model unobserved variables such as (rational) expectations, measurement errors, missing observations, permanent income, unobserved components (cycles and trends), and the natural rate of unemployment. The Kalman filter algorithm has been used to compute exact, finite sample forecasts for Gaussian ARMA models, multivariate (vector) ARMA models, MIMIC (multiple indicators and multiple causes), Markov switching models, and time varying (random) coefficient models. Extensive surveys of applications of state space models in econometrics can be found in Hamilton (1994a, chapter 13; 1994b) and Harvey (1989, chapters 3,

4). Many time-series models used in econometrics are special cases of the class of linear state space models developed by engineers to describe physical systems. The Kalman filter, an efficient recursive method for computing optimal linear forecasts in such models, can be exploited to compute the exact Gaussian likelihood function.

Eviews estimates both single and multiple equation dynamic models using a state space object. The state space object provides you with tools for specifying, estimating, and working with the results of your dynamic system. This Chapter is organized as follows. In the next section we present different forms of representation and estimation of state space models. In section 3, we present the Kalman Filter. Finally the empirical investigation is presented in section 4.

5.2 Representation and Estimation

We present here a brief discussion of the specification and estimation of a state space model. Those desiring greater detail are directed to Hamilton (1994a, chapter 13) and Harvey (1993, chapter 4). The state space representation of the dynamics of an $n \times 1$ vector Y is given by the following system of equations:

$$Y_t = A'_t x_t + F'_t \theta_t + v_t \quad (1)$$

$$\theta_{t+1} = G_t \theta_t + w_{t+1} \quad (2)$$

for where A', F', G are matrices of dimension $n \times q$, $n \times k$, and $k \times k$, respectively. x is a $q \times 1$ vector of exogenous or predetermined variables and θ is a $k \times 1$ vector of possibly

unobserved state variables. The first equation is known as the observation (or measurement) equation and the second is known as the state (or transition) equation. The disturbance vectors v_t and w_t are assumed to be independent white noise with

$$\text{var}(v_t) = R, \text{var}(w_t) = Q, E(v_t w_t) = 0$$

for all s and t . Given observations (Y_t, x_t) for $t = 1, 2, \dots, T$, our goal is to estimate the parameters A, F, G, R, Q , and make inferences about the state vector θ . The following broad classes of state space models are estimated with regards to this study:

- Recursive Coefficient Models.
- ARMA Models

Recursive Coefficient Models.

The class of models that can be estimated with this option have the following form:

$$Y_t = z_t' \beta_t + w_t$$

$$\beta_{t+1} = \beta_t$$

where the state vector is the recursive coefficient vector β_t . Recursive estimation restricts the state equation to have zero variance and unit coefficient. While the parameters in time-varying estimation are stochastic, the parameters in recursive estimation are deterministic.

ARMA Models

The class of models that can be estimated with this option have the following form:

$$Y_t = A'x_t + F'\xi_t$$

$$\xi_{t+1} = G\xi_t + v_{t+1}$$

ARMAX models (ARMA models with exogenous variables) are written in state space form by restricting the observation equation to have no error term. The state vector is part of the disturbance term and the shocks v in the state equation are the innovations. The coefficient matrix F in the observation equation are the MA parameters, while the coefficient matrix G in the state equation are the AR parameters. (Troskie (2001))

5.3 The Kalman Filter

5.3.1 Introduction

The Kalman filter is a recursive algorithm for sequentially updating the state vector given past information. More technically, it is an algorithm for calculating linear least squares forecasts of the state vector given data observed up to date t . The state vector θ and its mean squared error $P_t = E \left[\left(\theta_t - \hat{\theta}_t \right) \left(\theta_t - \hat{\theta}_t \right)' \right]$ are recursively estimated by:

$$\begin{aligned}\theta_{t|t} &= G\theta_{t-1|t-1} + S_{t-1}F(F'S_{t-1}F + R)^{-1}(Y_t - A'x_t - F'G\theta_{t-1|t-1}) \\ P_{t|t} &= S_{t-1} - S_{t-1}F(F'S_{t-1}F + R)^{-1}F'S_{t-1}\end{aligned}$$

where $S_{t-1} = FP_{t-1|t-1}F' + Q$, and $\theta_{s|v}$ is the forecast of the state vector at time period s , given information available at time v . Note that the recursion for P does not depend on the forecasted state vector $\theta_{t-1|t-1}$, or on the observed data (Y_t, x_t) . To implement the Kalman filter, we must specify the starting values and replace the unknown matrices by their estimates. By default, Eviews obtains starting values by treating these matrices as fixed coefficients and estimating them using ordinary least squares (OLS).

the vector of interest to be estimated and the vector \mathbf{x} is a $(q \times 1)$ vector of exogenous or predetermined variables. The equation (1) is called the observation or signal equation and is more often written as

$$\mathbf{Y}_t = \mathbf{F}'_t \boldsymbol{\theta}_t + \mathbf{v}_t \quad (5)$$

and equation (2) is called the state space (SS) equation. The error terms \mathbf{v}_t and \mathbf{w}_t are often assumed to be uncorrelated with respect to time t and also mutually uncorrelated ($\mathbf{H}_t = \mathbf{0}$). In the case of normality this implies independence and mutually independence. To derive the estimation algorithms called the Kalman Filter we assume, for convenience sake,

$$\mathbf{Y}_t = \mathbf{F}'_t \boldsymbol{\theta}_t + \mathbf{v}_t, \quad \mathbf{v}_t \sim N(\mathbf{0}, \mathbf{V}_t) \quad (6)$$

$$\boldsymbol{\theta}_t = \mathbf{G}_t \boldsymbol{\theta}_{t-1} + \mathbf{w}_t, \quad \mathbf{w}_t \sim N(\mathbf{0}, \mathbf{W}_t).$$

5.3.3 Recursive Estimation of the Model.

Let the model be

$$\mathbf{Y}_t = \mathbf{F}'_t \boldsymbol{\theta}_t + \mathbf{v}_t, \quad \mathbf{v}_t \sim N(\mathbf{0}, \mathbf{V}_t) \quad (23)$$

$$\boldsymbol{\theta}_t = \mathbf{G}_t \boldsymbol{\theta}_{t-1} + \mathbf{w}_t, \quad \mathbf{w}_t \sim N(\mathbf{0}, \mathbf{W}_t).$$

where the error components are serially independent and mutually independent. The prior information given at $t = 0$ is given by D_0 with a guess $\hat{\boldsymbol{\theta}}_0$ for the mean vector and a guess Σ_0 for the covariance matrix with a multivariate distribution for $\boldsymbol{\theta}_0$ given by

$$(\boldsymbol{\theta}_0/D_0) \sim N(\hat{\boldsymbol{\theta}}_0, \Sigma_0) \quad (24)$$

and independent of $\mathbf{v}_t, \mathbf{w}_t$. The information set available at time t is then simply $D_t = \{\mathbf{Y}_t, D_{t-1}\}$. The recursive procedure, starting at time $t = 0$ is then as follows. At each point t in time there are two stages

Stage I: Before \mathbf{Y}_t is observed and

Stage II After \mathbf{Y}_t is observed.

Stage I.

At time $t - 1$ the distribution of

$$(\boldsymbol{\theta}_{t-1}/D_{t-1}) \sim N(\hat{\boldsymbol{\theta}}_{t-1}, \boldsymbol{\Sigma}_{t-1})$$

for some $\hat{\boldsymbol{\theta}}_{t-1}$ and $\boldsymbol{\Sigma}_{t-1}$. Before \mathbf{Y}_t is observed the best estimate of the value of $\boldsymbol{\theta}_t$ can be obtained from the state space system of equations

$$\boldsymbol{\theta}_t = \mathbf{G}_t \boldsymbol{\theta}_{t-1} + \mathbf{w}_t \quad (25)$$

so that at time $t - 1$ given \mathbf{Y}_{t-1} the expected value of $\boldsymbol{\theta}_t$ is given by

$$\begin{aligned} E(\boldsymbol{\theta}_t/\mathbf{Y}_{t-1}) &= E(\mathbf{G}_t \boldsymbol{\theta}_{t-1}/\mathbf{Y}_{t-1}) + E(\mathbf{w}_t/\mathbf{Y}_{t-1}) \\ &= \mathbf{G}_t \hat{\boldsymbol{\theta}}_{t-1} \end{aligned} \quad (26)$$

and covariance matrix

$$\text{var}(\boldsymbol{\theta}_t/\mathbf{Y}_{t-1}) = \text{var}[(\mathbf{G}_t \boldsymbol{\theta}_{t-1} + \mathbf{w}_t)/\mathbf{Y}_{t-1}] \quad (27)$$

$$\begin{aligned} &= \text{var}(\mathbf{G}_t \boldsymbol{\theta}_{t-1}/\mathbf{Y}_{t-1}) + \text{var}(\mathbf{w}_t/\mathbf{Y}_{t-1}) \\ &= \mathbf{G}_t \boldsymbol{\Sigma}_{t-1} \mathbf{G}_t' + \mathbf{W}_t = \mathbf{R}_t \text{ (say).} \end{aligned} \quad (28)$$

Thus the distribution of $\boldsymbol{\theta}_t$ given \mathbf{Y}_{t-1} is

$$(\boldsymbol{\theta}_t/\mathbf{Y}_{t-1}) \sim N(\mathbf{G}_t \hat{\boldsymbol{\theta}}_{t-1}, \mathbf{R}_t). \quad (29)$$

The distribution (29) is the prior distribution of θ_t given Y_{t-1} . Thus before Y_t is observed, the estimate of θ_t is the expected value of (θ_t/Y_{t-1}) , that is

$$E(\theta_t/Y_{t-1}) = G_t \hat{\theta}_{t-1} = \hat{\theta}_{t/t-1} \quad (30)$$

and the (prior) prediction or forecast of Y_t (before it is observed) is then

$$\hat{Y}_{t/t-1} = F' \hat{\theta}_{t/t-1} = F' G_t \hat{\theta}_{t-1} \quad (31)$$

Stage II

After Y_t is observed we need to derive the posterior distribution of $(\theta_t/Y_t, D_{t-1})$.

Lemma 3.1

Let X be distributed like a multivariate normal distribution $X \sim N(\mu, \Sigma)$ and let X be partitioned as

$$X = \begin{pmatrix} X^{(1)} \\ X^{(2)} \end{pmatrix}$$

into q and $r = p - q$ components respectively. Partition μ and Σ conformably, that is

$$\mu = \begin{pmatrix} \mu^{(1)} \\ \mu^{(2)} \end{pmatrix} \text{ and } \Sigma = \begin{pmatrix} \Sigma_{11} & \Sigma_{12} \\ \Sigma_{21} & \Sigma_{22} \end{pmatrix}.$$

The conditional distribution of $X^{(1)}$ given $X^{(2)} = x^{(2)}$ is then

$$N(\mu^{(1)} + \Sigma_{12}\Sigma_{22}^{-1}(x^{(2)} - \mu^{(2)}), \Sigma_{11.2})$$

.which is a multivariate normal density with conditional mean

$$E(X^{(1)}/x^{(2)}) = \mu^{(1)} + \Sigma_{12}\Sigma_{22}^{-1}(x^{(2)} - \mu^{(2)})$$

and conditional covariance matrix

$$\text{cov}(X^{(1)}/x^{(2)}) = \Sigma_{11.2} = \Sigma_{11} - \Sigma_{12}\Sigma_{22}^{-1}\Sigma_{21}.$$

Note that the conditional mean $E(\mathbf{X}^{(1)}/\mathbf{x}^{(2)})$ depends on $\mathbf{x}^{(2)}$ but the conditional covariance matrix $\text{cov}(\mathbf{X}^{(1)}/\mathbf{x}^{(2)})$ does not depend on $\mathbf{x}^{(2)}$. From the conditional covariance matrix, also called the partial covariance matrix (Fisher),

$$\begin{aligned}\Sigma_{11.2} &= \Sigma_{11} - \Sigma_{12}\Sigma_{22}^{-1}\Sigma_{21} \\ &= [\sigma_{ij.q+1,\dots,p}], \quad i, j = 1, \dots, q\end{aligned}$$

we can also compute the partial (conditional) covariance matrix

$$[\rho_{ij.q=1,\dots,p}] = \left[\frac{\sigma_{ij.q+1,\dots,p}}{\sqrt{\sigma_{ii.q+1,\dots,p}\sigma_{jj.q+1,\dots,p}}} \right], \quad i, j = 1, \dots, q.$$

If the correlation matrix

$$\mathbf{P} = [\rho_{ij}], \quad i, j = 1, \dots, p$$

is partitioned conformably to Σ , as

$$\mathbf{P} = \begin{pmatrix} \mathbf{P}_{11} & \mathbf{P}_{12} \\ \mathbf{P}_{21} & \mathbf{P}_{22} \end{pmatrix}$$

then the matrix

$$\begin{aligned}\mathbf{P}_{11.2} &= \mathbf{P}_{11} - \mathbf{P}_{12}\mathbf{P}_{22}^{-1}\mathbf{P}_{21} \\ &= [\rho_{ij.q=1,\dots,p}], \quad i, j = 1, \dots, q.\end{aligned}$$

is called the partial correlation matrix.

Stage II (continued)

We have observed \mathbf{Y}_t and want to compute posterior distribution of $(\boldsymbol{\theta}_t/\mathbf{Y})_t, D_{t-1}$.

Let \mathbf{e}_t be the error in the prediction of \mathbf{Y}_t at time $t - 1$, that is

$$\begin{aligned}\mathbf{e}_t &= \mathbf{Y}_t - \hat{\mathbf{Y}}_t \\ &= \mathbf{F}'_t \boldsymbol{\theta}_t + \mathbf{v}_t - \mathbf{F}'_t \mathbf{G}_t \hat{\boldsymbol{\theta}}_{t-1} \\ &= \mathbf{F}'_t (\boldsymbol{\theta}_t - \mathbf{G}_t \hat{\boldsymbol{\theta}}_{t-1}) + \mathbf{v}_t\end{aligned}\tag{31}$$

Since $\mathbf{v}_t \sim N(\mathbf{0}, \mathbf{V}_t)$ independently from $\boldsymbol{\theta}_t$ the conditional distribution of \mathbf{e}_t given $\boldsymbol{\theta}_t$ and of course \mathbf{Y}_{t-1} is

$$(\mathbf{e}_t/\boldsymbol{\theta}_t, \mathbf{Y}_{t-1}) \sim N[\mathbf{F}'_t(\boldsymbol{\theta}_t - \mathbf{G}_t \hat{\boldsymbol{\theta}}_{t-1}), \mathbf{V}_t].\tag{32}$$

Note that $\mathbf{F}'_t, \mathbf{G}_t$ and \mathbf{V}_t are known, it follows that to observe \mathbf{e}_t is the same as to observe \mathbf{Y}_t . Suppose now in Lemma 3.1

$$\mathbf{X}^{(1)} = (\mathbf{e}_t/\mathbf{Y}_{t-1}) \sim N(\boldsymbol{\mu}^{(1)}, \boldsymbol{\Sigma}_{11}),$$

where we wish to determine $\boldsymbol{\mu}^{(1)}$ and $\boldsymbol{\Sigma}_{11}$ and

$$\mathbf{X}^{(2)} = (\boldsymbol{\theta}_t/\mathbf{Y}_{t-1}) \sim N(\boldsymbol{\mu}^{(2)}, \boldsymbol{\Sigma}_{22}) \sim N(\mathbf{G}_t \hat{\boldsymbol{\theta}}_{t-1}, \mathbf{R}_t) \text{ from (29).}$$

Then $\boldsymbol{\mu}^{(2)} = \mathbf{G}_t \hat{\boldsymbol{\theta}}_{t-1}$ and $\boldsymbol{\Sigma}_{22} = \mathbf{R}_t$. Now consider (32) in terms of Lemma 3.1.

$$(\mathbf{X}^{(1)}/\mathbf{X}^{(2)}) \sim N(\boldsymbol{\mu}^{(1)} + \boldsymbol{\Sigma}_{12} \boldsymbol{\Sigma}_{22}^{-1}(\mathbf{x}^{(2)} - \boldsymbol{\mu}^{(2)}), \boldsymbol{\Sigma}_{11} - \boldsymbol{\Sigma}_{12} \boldsymbol{\Sigma}_{22}^{-1} \boldsymbol{\Sigma}_{21}),\tag{33}$$

$$\text{where } (\mathbf{e}_t/\boldsymbol{\theta}_t, \mathbf{Y}_{t-1}) = (\mathbf{X}^{(1)}/\mathbf{X}^{(2)})$$

so that

$$\boldsymbol{\mu}^{(1)} + \boldsymbol{\Sigma}_{12} \boldsymbol{\Sigma}_{22}^{-1}(\mathbf{x}^{(2)} - \boldsymbol{\mu}^{(2)}) = \mathbf{F}'_t(\boldsymbol{\theta}_t - \mathbf{G}_t \hat{\boldsymbol{\theta}}_{t-1})$$

i.e.

$$\boldsymbol{\mu}^{(1)} + \boldsymbol{\Sigma}_{12} \mathbf{R}_t^{-1}(\boldsymbol{\theta}_t - \mathbf{G}_t \hat{\boldsymbol{\theta}}_{t-1}) = \mathbf{F}'_t(\boldsymbol{\theta}_t - \mathbf{G}_t \hat{\boldsymbol{\theta}}_{t-1}).$$

Thus

$$\mu^{(1)} = 0 ; \Sigma_{12} R_t^{-1} = F_t' \text{ and } \Sigma_{12} = F_t' R_t.$$

Comparing the covariance of (32) and (33) we see that

$$\Sigma_{11} - \Sigma_{12} \Sigma_{22}^{-1} \Sigma_{21} = V_t$$

$$\Sigma_{11} - F_t' R_t R_t^{-1} R_t F_t = V_t$$

yielding

$$\Sigma_{11} = V_t + F_t' R_t F_t.$$

Thus

$$X^{(1)} \cong (e_t / Y_{t-1}) \sim N(0, (V_t + F_t' R_t F_t))$$

and jointly

$$X = \begin{pmatrix} X^{(1)} \\ X^{(2)} \end{pmatrix} \sim N \left[\begin{pmatrix} \mu^{(1)} \\ \mu^{(2)} \end{pmatrix}, \begin{pmatrix} \Sigma_{11} & \Sigma_{12} \\ \Sigma_{21} & \Sigma_{22} \end{pmatrix} \right]$$

that is

$$\begin{pmatrix} e_t \\ \theta_t \end{pmatrix} / Y_{t-1} \sim N \left(\begin{bmatrix} 0 \\ G_t \hat{\theta}_{t-1} \end{bmatrix}, \begin{bmatrix} V_t + F_t' R_t F_t & F_t' R_t \\ R_t F_t & R_t \end{bmatrix} \right) \quad (34)$$

Consequently from Lemma 3.1

$$(X^{(2)} / X^{(1)}) \sim N(\mu^{(2)} + \Sigma_{21} \Sigma_{11}^{-1} (x^{(1)} - \mu^{(1)}), \Sigma_{22} - \Sigma_{21} \Sigma_{11}^{-1} \Sigma_{12})$$

yields the posterior distribution of

$$(\theta_t / e_t, Y_{t-1}) = (\theta_t / Y_t)$$

$$\sim N(G_t \hat{\theta}_{t-1} + R_t F_t (V_t + F_t' R_t F_t)^{-1} e_t, R_t - R_t F_t (V_t + F_t' R_t F_t)^{-1} F_t' R_t)$$

Thus the posterior estimation of θ_t is:

$$\hat{\theta}_t = E(\theta_t/Y_t) \quad (35)$$

$$= G_t \hat{\theta}_{t-1} + R_t F_t (V_t + F_t' R_t F_t)^{-1} e_t,$$

$$\text{where } e_t = Y_t - \hat{Y}_t = Y_t - F_t' G_t \hat{\theta}_{t-1}, \quad (36)$$

$$R_t = G_t \Sigma_{t-1} G_t' + W_t, \text{ and} \quad (37)$$

$$\text{var}(\theta_t/Y_t) = \Sigma_t \quad (38)$$

$$= R_t - R_t F_t (V_t + F_t' R_t F_t)^{-1} F_t' R_t$$

Remark: For the general formulation given by equations (1) and (2) use the following replacements

$$\begin{aligned} e_t &= Y_t - \hat{Y}_{t/t-1} \\ &= Y_t - c_t - A x_t - d_t - F_t' G_t \hat{\theta}_{t-1}, \end{aligned} \quad (39)$$

and

$$V_t \Rightarrow S_t V_t S_t', \quad W_t \Rightarrow K_t W_t K_t'.$$

For computing e_t given by (39) we have assumed that all the other parameters (c_t , d_t , A , G_t , V_t and W_t) are deterministic. In practice they have to be estimated from the data $\{Y_t, x_t, t = 1, \dots, T\}$.

5.3.4 The Kalman Filtering Process.

The filtering process begins at time $t = 0$ where we have prior knowledge about

$$(\theta_0/D_0) \sim N(\hat{\theta}_0, \Sigma_0)$$

which is independent of \mathbf{v}_t and \mathbf{w}_t . At times $t = 1, 2, \dots, T$ we use the filtering equations (35) to (38) to obtain at time t

$$\hat{\boldsymbol{\theta}}_t \text{ and } \boldsymbol{\Sigma}_t$$

5.3.5 Residuals and Maximum Likelihood Estimates.

Assuming that the signal (observation) model and state space model are given by equations (1) and (2) (in a modified, but more convenient, form)

$$\mathbf{Y}_t = \mathbf{A}'_t \mathbf{x}_t + \mathbf{F}'_t \boldsymbol{\theta}_t + \mathbf{v}_t, \quad \mathbf{v}_t \sim N(\mathbf{0}, \mathbf{V}_t), \quad (40)$$

$$\boldsymbol{\theta}_t = \mathbf{G}_t \boldsymbol{\theta}_{t-1} + \mathbf{w}_t, \quad \mathbf{w}_t \sim N(\mathbf{0}, \mathbf{W}_t), \quad (41)$$

the data observed at time t is given by $(\mathbf{Y}_t, \mathbf{x}_t)$. The residuals are computed from the signal (observation) equation as

$$\mathbf{e}_t = \mathbf{Y}_t - \mathbf{A}'_t \mathbf{x}_t - \mathbf{F}'_t \mathbf{G}_t \hat{\boldsymbol{\theta}}_{t-1} \quad (42)$$

see also (31) and are the one-step ahead residuals. Observing \mathbf{e}_t is the same as observing \mathbf{Y}_t . The distribution of \mathbf{e}_t is given by (34) as

$$(\mathbf{e}_t / \mathbf{Y}_{t-1}) \sim N[\mathbf{0}, \mathbf{V}_t + \mathbf{F}'_t \mathbf{R}_t \mathbf{F}_t]. \quad (43)$$

that is

$$\begin{aligned} f(\mathbf{e}_t / \mathbf{x}_t, \mathbf{Y}_{t-1}) &= (2\pi)^{-n/2} \left| \mathbf{V}_t + \mathbf{F}'_t \mathbf{R}_t \mathbf{F}_t \right|^{-1/2} \\ &\times \exp[-1/2 \{ \mathbf{e}_t (\mathbf{V}_t + \mathbf{F}'_t \mathbf{R}_t \mathbf{F}_t)^{-1} \mathbf{e}_t \}], \end{aligned} \quad (44)$$

for $t = 1, 2, \dots, T$. The log likelihood is therefore

$$\sum_{t=1}^T \log f(\mathbf{e}_t / \mathbf{x}_t, \mathbf{Y}_{t-1}). \quad (45)$$

Expression (45) can then be maximized numerically with respect to the unknown parameters in

$$A, F, G, V \text{ and } W$$

This usually involves complicated non-linear numerical optimization. The computation is simplified when only $ARMA(p, q)$ parameters are involved. We now complete this study with an empirical investigation. (Troskie (2001))

5.4 Empirical Investigation: Single Index

5.4.1 Introduction

The early development of state space methodology took place in the field of engineering rather than statistics, starting with the breaking paper of Kalman (1960). In this paper Kalman did two crucially important things. He showed that very wide class of problems could be encapsulated in a simple linear model, essentially the state space model. Secondly he showed how, due to Markovian nature of the model, the calculations needed for practical application of the model could be set up in recursive form in a way that was particularly convenient for computers.

A colossal load of work was achieved in the development of these ideas in the engineering field subsequently. In the 1960s to early 1980's contributions to state space methodology from statisticians and econometricians were isolated and sporadic. In recent years however there has been rapid growth of interest in the field of both statistics and econometrics as is indicated by the empirical studies conducted thus far. We now commence the

next study, which is to establish single and multiple index state space models for the nine financial stocks in our portfolio.

5.4.2 The data

For this empirical analysis, the data used is the SIDS batch.

5.4.3 Study Objectives

Having established single index time series and regression models for $R_1, R_2, R_3, R_4, R_5, R_6, R_7, R_8$ and R_9 , our next objective is to produce state space models for each of these nine response variables. Furthermore by means of an empirical check, comparisons will be conducted in order to establish that state space modelling is far superior to time series modelling. The superiority objective will be further explored in the portfolio theory chapter to follow with regards to mean-variance frontiers. For now we shall construct state space models for the 9 responses and compare the models' efficiencies with that of their time series counterpart.

5.4.4 Methodology

Referring to section 5.3.3 of this chapter, 'Representation and Estimation', it was mentioned there, that the following broad classes of state space models are estimated with regards to this study:

- Recursive Coefficient Models
- and ARMA (autoregressive moving average) models.

The ARMA state space model is only going to be fitted to those responses that exhibited ARMA components from the chapter 4, namely for responses variables R_1 , R_3 , R_4 , R_8 and R_9 . Thus the Recursive Coefficient Model will be fitted to the remaining 4. The time series package Eviews 3.1 is implemented for all model formulations and model selection criteria calculations.

Once fitting the respective models for all 9 responses, the model selection procedure is based on the R^2 , Adjusted R^2 , SE of Regression and the Durbin Watson (D-W) Statistic. Unfortunately, Eviews 3.1 does not facilitate for the calculation for the AIC and BIC of state space objects. Once the appropriate state space models are fit for the respective responses, the most efficient model is opted for based on the aforementioned selection criteria. Accompanying the selection procedure we will continuously provide a comparison between state space and time series modelling in terms of model fit.

5.4.5 Primary Findings

From the previous empirical study, time series models were established for the nine financial stocks in our portfolio. The above outlined methodology was applied for each of the time series models acquired from the previous chapter in the appropriate state space form. We highlight the results from applying the state space estimation procedures for our portfolio. Table 5.1 below illustrates a summary of the model selection criteria from fitting the new models

State Space Models	Model Selection Criteria		
	Adj- R^2	S.E. of reg	D-W stat
Model for R1 (Model 1)	0.6893	0.0557	2.0343
Model for R2 (Model 2)	0.2016	0.1214	2.1542
Model for R3 (Model 3)	0.2222	0.0901	1.9716
Model for R4 (Model 4)	0.4163	0.0608	1.9548
Model for R5 (Model 5)	0.0540	0.0819	1.9696
Model for R6 (Model 6)	0.3338	0.0907	1.8827
Model for R7 (Model 7)	0.3716	0.0779	1.8905
Model for R8 (Model 8)	0.2977	0.0652	2.1353
Model for R9 (Model 9)	0.2375	0.0802	2.0382

table 5.1: summary of the model selection criteria

Before any further analysis of the results, table 5.2 below presents the state space forms adopted for each of the responses variables R_1 through to R_9 .

State Space Models	Definition Of State Space Structure
Model for R1 (Model 1)	$R1 = C(1) + C(2)*JSE + SV1$ $SV1 = C(3)*SV1(-2)$
Model for R2 (Model 2)	$R2 = SV1 + SV2*JSE$ $SV1 = SV1(-1)$ $SV2 = SV2(-1)$
Model for R3 (Model 3)	$R3 = C(1) + C(2)*JSE + SV1$ $SV1 = C(3)*SV1(-1)$
Model for R4 (Model 4)	$R4 = C(1) + C(2)*JSE + SV1$ $SV1 = C(3)*SV1(-1)$
Model for R5 (Model 5)	$R5 = SV1 + SV2*JSE$ $SV1 = SV1(-1)$ $SV2 = SV2(-1)$
Model for R6 (Model 6)	$R6 = SV1 + SV2*JSE$ $SV1 = SV1(-1)$ $SV2 = SV2(-1)$
Model for R7 (Model 7)	$R7 = SV1 + SV2*JSE$ $SV1 = SV1(-1)$ $SV2 = SV2(-1)$
Model for R8 (Model 8)	$R8 = C(1) + C(2)*JSE + SV1$ $SV1 = C(3)*SV1(-6)$
Model for R9 (Model 9)	$R9 = C(1) + C(2)*JSE + SV1$ $SV1 = C(3)*SV1(-2)$

Table 5.2: State Space form of Stocks

Examining the results from table 5.1 above, it is blatantly clear that all nine $Adj - R^2$ statistics have increased thus explaining more variation than their time series counterparts. Accompanying the higher achieved $Adj - R^2$, the state space models have also successfully reduced the SE of Regression for all nine responses. Thus the state space models provide better explanatory power in terms of higher R^2 and lower error variances of regression. This implies that the variation in the security returns that is explained by the variation

in the JSE-all share index is increased in the state space scenario. The lower error variances are directly the result of the filtering technique, which serves to smooth noise and strike out variability. The effect of lower error variances is going to contribute to the variance and covariance structure of the stocks during the portfolio optimisation and efficient frontier construction phases. Next we examine the significance of the new updated coefficients. First we examine the coefficients of the ARMA state space models presented in table 5.3 below.

ARMA State Space Model Coefficients				
ARMA State Space Models	Coefficient of JSE C(2)			
	C(2)	Std. Error	t-Statistic	Prob.
Model for R1 (Model 1)	1.4199	0.0346	41.0482	0.0000
Model for R3 (Model 3)	0.7044	0.1169	6.0248	0.0000
Model for R4 (Model 4)	0.8199	0.0821	9.9883	0.0000
Model for R8 (Model 8)	0.7340	0.0740	9.9242	0.0000
Model for R9 (Model 9)	0.7460	0.1059	7.0460	0.0000
ARMA State Space Models	Coefficient of AR(p) C(3)			
	C(3)	Std. Error	t-Statistic	Prob.
Model for R1 (Model 1)	-0.1386	0.0605	-2.2897	0.0231
Model for R3 (Model 3)	-0.2927	0.0593	-4.9363	0.0000
Model for R4 (Model 4)	-0.1527	0.0621	-2.4575	0.0149
Model for R8 (Model 8)	-0.2360	0.0658	-3.5886	0.0004
Model for R9 (Model 9)	-0.1864	0.0572	-3.2601	0.0013

Table 5.3: ARMA state space model coefficients

It is clear from table 5.3 above that the coefficients are all significant at the 5% level of significance. Thus the ARMA state space models are all significant. Table 5.4 below summarises the coefficients of the recursive coefficient state space models. A similar scenario is observed for the coefficients of the recursive state space method. All coefficients are significant at the 5% level of significance.

Recursive Coefficient State Space Model Coefficients				
Recursive Coefficient Models	Coefficient of JSE SV2			
	Final SV2	Std. Error	t-Statistic	Prob.
Model for R2 (Model 2)	1.0359	0.1478	7.0083	0.0000
Model for R5 (Model 5)	0.3115	0.0985	3.1606	0.0018
Model for R6 (Model 6)	1.0776	0.1092	9.8721	0.0000
Model for R7 (Model 7)	1.0141	0.0944	10.7397	0.0000

Table 5.4: Recursive state space model coefficients

State Space Models	Model Selection Criteria		
	Adj- R^2	S.E. of reg	D-W stat
Model for R1 (Model 1)	0.6893	0.0557	2.0343
Model for R2 (Model 2)	0.2016	0.1214	2.1542
Model for R3 (Model 3)	0.2222	0.0901	1.9716
Model for R4 (Model 4)	0.4163	0.0608	1.9548
Model for R5 (Model 5)	0.0540	0.0819	1.9696
Model for R6 (Model 6)	0.3338	0.0907	1.8827
Model for R7 (Model 7)	0.3716	0.0779	1.8905
Model for R8 (Model 8)	0.2977	0.0652	2.1353
Model for R9 (Model 9)	0.2375	0.0802	2.0382

table 5.1: summary of the model selection criteria

Before any further analysis of the results, table 5.2 below presents the state space forms adopted for each of the responses variables R_1 through to R_9 .

State Space Models	Definition Of State Space Structure
Model for R1 (Model 1)	$R_1 = C(1) + C(2)*JSE + SV1$ $SV1 = C(3)*SV1(-2)$
Model for R2 (Model 2)	$R_2 = SV1 + SV2*JSE$ $SV1 = SV1(-1)$ $SV2 = SV2(-1)$
Model for R3 (Model 3)	$R_3 = C(1) + C(2)*JSE + SV1$ $SV1 = C(3)*SV1(-1)$
Model for R4 (Model 4)	$R_4 = C(1) + C(2)*JSE + SV1$ $SV1 = C(3)*SV1(-1)$
Model for R5 (Model 5)	$R_5 = SV1 + SV2*JSE$ $SV1 = SV1(-1)$ $SV2 = SV2(-1)$
Model for R6 (Model 6)	$R_6 = SV1 + SV2*JSE$ $SV1 = SV1(-1)$ $SV2 = SV2(-1)$
Model for R7 (Model 7)	$R_7 = SV1 + SV2*JSE$ $SV1 = SV1(-1)$ $SV2 = SV2(-1)$
Model for R8 (Model 8)	$R_8 = C(1) + C(2)*JSE + SV1$ $SV1 = C(3)*SV1(-6)$
Model for R9 (Model 9)	$R_9 = C(1) + C(2)*JSE + SV1$ $SV1 = C(3)*SV1(-2)$

Table 5.2: State Space form of Stocks

Examining the results from table 5.1 above, it is blatantly clear that all nine $Adj - R^2$ statistics have increased thus explaining more variation than their time series counterparts. Accompanying the higher achieved $Adj - R^2$, the state space models have also successfully reduced the SE of Regression for all nine responses. Thus the state space models provide better explanatory power in terms of higher R^2 and lower error variances of regression. This implies that the variation in the security returns that is explained by the variation

in the JSE-all share index is increased in the state space scenario. The lower error variances are directly the result of the filtering technique, which serves to smooth noise and strike out variability. The effect of lower error variances is going to contribute to the variance and covariance structure of the stocks during the portfolio optimisation and efficient frontier construction phases. Next we examine the significance of the new updated coefficients. First we examine the coefficients of the ARMA state space models presented in table 5.3 below.

ARMA State Space Model Coefficients				
ARMA State Space Models	Coefficient of JSE C(2)			
	C(2)	Std. Error	t-Statistic	Prob.
Model for R1 (Model 1)	1.4199	0.0346	41.0482	0.0000
Model for R3 (Model 3)	0.7044	0.1169	6.0248	0.0000
Model for R4 (Model 4)	0.8199	0.0821	9.9883	0.0000
Model for R8 (Model 8)	0.7340	0.0740	9.9242	0.0000
Model for R9 (Model 9)	0.7460	0.1059	7.0460	0.0000
ARMA State Space Models	Coefficient of AR(p) C(3)			
	C(3)	Std. Error	t-Statistic	Prob.
Model for R1 (Model 1)	-0.1386	0.0605	-2.2897	0.0231
Model for R3 (Model 3)	-0.2927	0.0593	-4.9363	0.0000
Model for R4 (Model 4)	-0.1527	0.0621	-2.4575	0.0149
Model for R8 (Model 8)	-0.2360	0.0658	-3.5886	0.0004
Model for R9 (Model 9)	-0.1864	0.0572	-3.2601	0.0013

Table 5.3: ARMA state space model coefficients

It is clear from table 5.3 above that the coefficients are all significant at the 5% level of significance. Thus the ARMA state space models are all significant. Table 5.4 below summarises the coefficients of the recursive coefficient state space models. A similar scenario is observed for the coefficients of the recursive state space method. All coefficients are significant at the 5% level of significance.

Recursive Coefficient State Space Model Coefficients				
Recursive Coefficient Models	Coefficient of JSE SV2			
	Final SV2	Std. Error	t-Statistic	Prob.
Model for R2 (Model 2)	1.0359	0.1478	7.0083	0.0000
Model for R5 (Model 5)	0.3115	0.0985	3.1606	0.0018
Model for R6 (Model 6)	1.0776	0.1092	9.8721	0.0000
Model for R7 (Model 7)	1.0141	0.0944	10.7397	0.0000

Table 5.4: Recursive state space model coefficients

The conclusions follow regarding the above findings in the subsequent section.

5.4.6 Conclusions

Having completed implementing the state space methodologies on the respective responses, the following is concluded: that significant state space models have been found for all nine stocks of our portfolio. Furthermore the state space models acquired appear to provide better explanatory power in terms of high R^2 values and low error variances of the models than the time series models for the stocks. Furthermore we conclude that State Space models are by far the most superior to both Time series and Regression Models as was illustrated by the findings section above. The betas and regression SE's computed for the nine stocks will be used as inputs for the computation of the efficient frontier for the single index state space model. The variance JSE and the mean log returns of the nine stocks will also be required as inputs for the computation of the state space (Kalman filter) method's mean-variance frontier.

The next empirical investigation considers the construction of state space models for the same 9 responses in a multiple index context.

5.5 Empirical Investigation: Multiple Index

5.5.1 The data

For this empirical analysis, the data used was the MIDS batch.

5.5.2 Study Objectives

Previously, we established single index state space (SISS) models for the 9 responses. We also showed that state space modelling is superior to time series and regression modelling in a single index environment. We try to achieve the same objectives in the multiple index setting.

5.5.3 Methodology

Recall the following broad classes of state space models:

- Recursive Coefficient Models
- and ARMA (autoregressive moving average) models.

In chapter 4, the response variables R_1, R_3, R_4, R_8 and R_9 exhibited the same time series components in both the single index (SI) and multiple index (MI) settings, hence the ARMA state space model is only going to be fitted again to these responses for the MI setting. Thus the Recursive Coefficient Model will be fitted to the remaining 4. The time series package Eviews 3.1 is implemented for all model formulations and model selection criteria calculations.

Once fitting the respective models for all 9 responses, the model selection procedure is based on the R^2 , Adjusted R^2 , SE of Regression and the Durbin Watson (D-W) Statistic. Unfortunately, Eviews 3.1 does not facilitate for the calculation for the AIC and BIC of state space objects. Once the appropriate state space models are fit for the respective responses, the most efficient model is opted for based on the aforementioned selection criteria. Accompanying the selection procedure we will continuously provide a comparison between MI state space and MI- time series modelling in terms of model fit.

5.5.4 Primary Findings

Implementing the above methodology to the respective responses, the following results are summarised in table 5.5 below:

State Space Models	Model Selection Criteria		
	Adj- R^2	S.E. of reg	D-W stat
Model for R1 (Model 1)	0.7127	0.0535	2.0380
Model for R2 (Model 2)	0.2244	0.1194	2.2267
Model for R3 (Model 3)	0.3063	0.0851	1.9833
Model for R4 (Model 4)	0.4198	0.0606	1.9545
Model for R5 (Model 5)	0.0753	0.0814	2.0503
Model for R6 (Model 6)	0.3899	0.0872	1.9141
Model for R7 (Model 7)	0.3908	0.0770	1.8511
Model for R8 (Model 8)	0.3399	0.0632	2.0864
Model for R9 (Model 9)	0.2604	0.0790	2.0564

Table 5.5: Summary of Selection criteria

Once again, the MI-state space models have consistently achieved higher $Adj - R^2$ and lower SEs of regression. Thus the state space techniques attained better explanatory power than their time series and regression complements. Furthermore the State space techniques provide better fit of the data than the other two modelling paradigms. The representations of the MI-state space models are given below in table 5.6:

State Space Models	Definition Of State Space Structure
Model for R1 (Model 1)	$R1 = C(1) + C(2)*X5 + C(3)*X6 + C(4)*X7 + SV1$ $SV1 = C(5)*SV1(-2)$
Model for R2 (Model 2)	$R2 = SV1_R2 + SV2_R2*X5 + SV3_R2*X6 + SV4_R2*X7$ $SV1_R2 = SV1_R2(-1), SV2_R2 = SV2_R2(-1)$ $SV3_R2 = SV3_R2(-1), SV4_R2 = SV4_R2(-1)$
Model for R3 (Model 3)	$R3 = C(1) + C(2)*X5 + C(3)*X6 + C(4)*X7 + SV1$ $SV1 = C(5)*SV1(-1)$
Model for R4 (Model 4)	$R4 = C(1) + C(2)*X5 + C(3)*X6 + C(4)*X7 + SV1$ $SV1 = C(5)*SV1(-1)$
Model for R5 (Model 5)	$R5 = SV1_R5 + SV2_R5*X5 + SV3_R5*X6 + SV4_R5*X7$ $SV1_R5 = SV1_R5(-1), SV2_R5 = SV2_R5(-1)$ $SV3_R5 = SV3_R5(-1), SV4_R5 = SV4_R5(-1)$
Model for R6 (Model 6)	$R6 = SV1_R6 + SV2_R6*X5 + SV3_R6*X6 + SV4_R6*X7$ $SV1_R6 = SV1_R6(-1), SV2_R6 = SV2_R6(-1)$ $SV3_R6 = SV3_R6(-1), SV4_R6 = SV4_R6(-1)$
Model for R7 (Model 7)	$R7 = SV1_R7 + SV2_R7*X5 + SV3_R7*X6 + SV4_R7*X7$ $SV1_R7 = SV1_R7(-1), SV2_R7 = SV2_R7(-1)$ $SV3_R7 = SV3_R7(-1), SV4_R7 = SV4_R7(-1)$
Model for R8 (Model 8)	$R8 = C(1) + C(2)*X5 + C(3)*X6 + C(4)*X7 + SV1$ $SV1 = C(5)*SV1(-6)$
Model for R9 (Model 9)	$R9 = C(1) + C(2)*X5 + C(3)*X6 + C(4)*X7 + SV1$ $SV1 = C(5)*SV1(-2)$

Table 5.6: MI-state space representations

Next we examine the significance of the new updated coefficients. The reader is referred to appendix 3A for the relevant statistics regarding the MI-state space coefficients. Scrutinising the time series coefficients for R_1, R_3, R_4, R_8 and R_9 , it is clear that they are all significant at the 5% level of significance. As seen before with our MI-time series and regression models, the subset of explanatory variables (X_5, X_6, X_7) is not always significant due to the problem of fitting on a common platform. Thus again it is clear in the state space scenario, that all the models do not always exhibit highly significant coefficients for this subset all the time. Nonetheless, the coefficients attained for this subset is non-zero, which is important for our mean-variance construction.

5.5.5 Conclusions

Firstly, significant MI-state space models were found for all 9 responses. In addition it can be concluded with confidence that the MI-State Space models more superior, in a mod-

elling context, to both Time series and Regression Models as was illustrated by the findings section above. The scenario is now set for the construction of State Space efficient frontiers for both the single and multiple index frameworks.

The next chapter looks at Volatility Modelling of stock returns by implementing GARCH models, a fairly new form of modelling. The previously found regression and time series models' volatilities will now be modelled using GARCH methods. This will be explored in the Empirical Study section of the subsequent chapter.

5.6 List of Appendices

5.6.1 Appendix 3A: MI-state space coefficients

	Coefficient	Std. Error	t-Statistic	Prob.
C(1)	-0.002439	0.003367	-0.724337	0.4697
C(2)	0.137862	0.035064	3.931721	0.0001
C(3)	1.225730	0.047584	25.75949	0.0000
C(4)	0.055358	0.030105	1.838812	0.0675
C(5)	-0.171551	0.068877	-2.490663	0.0136

R1 coefficient statistics

	Coefficient	Std. Error	t-Statistic	Prob.
C(1)	0.007389	0.005396	1.369245	0.1725
C(2)	-0.282562	0.052512	-5.380867	0.0000
C(3)	1.004987	0.124478	8.073602	0.0000
C(4)	0.058011	0.061200	0.947895	0.3444
C(5)	-0.283611	0.061888	-4.582627	0.0000

R3 coefficient statistics

	Coefficient	Std. Error	t-Statistic	Prob.
Final SV1_R5	0.009710	0.005869	1.654546	0.0997
Final SV2_R5	-0.114772	0.054953	-2.088559	0.0381
Final SV3_R5	0.459024	0.121301	3.784160	0.0002
Final SV4_R5	-0.012172	0.053964	-0.225565	0.8218

R5 coefficient statistics

	Coefficient	Std. Error	t-Statistic	Prob.
Final SV1_R7	0.005252	0.005598	0.938197	0.3493
Final SV2_R7	-0.051944	0.052414	-0.991039	0.3229
Final SV3_R7	1.012362	0.115698	8.750051	0.0000
Final SV4_R7	0.122534	0.051471	2.380623	0.0183

R7 coefficient statistics

	Coefficient	Std. Error	t-Statistic	Prob.
C(1)	0.001216	0.005049	0.240803	0.8100
C(2)	-0.043129	0.048452	-0.890139	0.3745
C(3)	0.721498	0.127869	5.642480	0.0000
C(4)	0.140128	0.042877	3.268151	0.0013
C(5)	-0.194176	0.059412	-3.268290	0.0013

R9 coefficient statistics

	Coefficient	Std. Error	t-Statistic	Prob.
Final SV1_R2	0.004717	0.008772	0.537719	0.5914
Final SV2_R2	-0.190363	0.082141	-2.317520	0.0215
Final SV3_R2	1.294221	0.181317	7.137908	0.0000
Final SV4_R2	-0.045490	0.080664	-0.563949	0.5735

R2 coefficient statistics

	Coefficient	Std. Error	t-Statistic	Prob.
C(1)	0.007869	0.004338	1.813817	0.0713
C(2)	-0.069103	0.038502	-1.794801	0.0743
C(3)	0.913789	0.094988	9.620035	0.0000
C(4)	-0.019330	0.034170	-0.565690	0.5723
C(5)	-0.159944	0.062526	-2.558031	0.0113

R4 coefficient statistics

	Coefficient	Std. Error	t-Statistic	Prob.
Final SV1_R6	-0.004430	0.006295	-0.703691	0.4825
Final SV2_R6	-0.014977	0.058945	-0.254081	0.7997
Final SV3_R6	0.965444	0.130113	7.420021	0.0000
Final SV4_R6	0.243997	0.057885	4.215231	0.0000

R6 coefficient statistics

	Coefficient	Std. Error	t-Statistic	Prob.
C(1)	0.003108	0.003745	0.829943	0.4076
C(2)	-0.154493	0.039544	-3.906871	0.0001
C(3)	0.928886	0.081325	11.42189	0.0000
C(4)	-0.029631	0.042468	-0.697729	0.4862
C(5)	-0.232860	0.069758	-3.338092	0.0010

R8 coefficient statistics

Chapter 6

Conditional Heteroskedastic Models-Volatility Models

6.1 Introduction

Risk as measured by variance plays a major role in many financial applications. The CAPM or Market Model (Sharpe (1964), Lintner (1965), Mossin(1966) and Merton(1973)) for example posits a positive relationship between the expected returns of a share and the market portfolio. This hypothesis relies on the assumption of a constant market variance. Portfolio managers as well as traders thus require an estimate of the level of future market uncertainty (variance) if they believe in the validity of the Capital Asset Pricing Model (CAPM). Volatility Models, first developed by Engle (1982), model the conditional volatility (conditional variance) of time series.

Volatility models have become increasingly important since asset return variances as well as the covariances of asset returns were found not to be constant over time (Bollerslev (1986)) but rather that they evolve over time. Many researchers have used volatility models in order to model share returns, exchange rate movements as well as interest rate movements. Both univariate time series models as well as their multivariate extensions are important to portfolio managers. For example, since covariances are time varying one could use a multivariate volatility model in order to solve the asset allocation problem of Markowitz (1952). Volatility models have also been used in order to test certain economic

theories. Bollerslev, Engle and Wooldridge (1988) used a trivariate CAPM in order to test the CAPM and the assumption of constant covariances and variances amongst the three assets used (treasury bills, bonds and stocks). They found that the conditional covariances between the three assets were varied over time. It was also shown that this movement could be forecasted.

Currently volatility models are most often used by derivatives analysts since they rely on estimates of the variance (vols) of certain financial instruments in order to use these estimates as inputs into their asset pricing models. Banks and large corporates however also use volatility models in order to estimate their daily value at risk (VAR). From a statistical point of view volatility models are important since many commonly used tools are no longer valid in the presence of non-constant variances. For example, standard regression type models assume that the residuals from the regression model is homoskedastic (constant over time). If this assumption is violated the analyst should adjust any results from such a model in order to compensate for heteroskedastic errors. ARCH in itself does not invalidate standard OLS inference. However, ignoring ARCH effects may result in the loss of efficiency of the estimated beta coefficients.

From the above discussion it can be seen that volatility models are important to both the financial practitioner as well as to the practicing statistician. In what follows is a discussion on both the theory and the use of volatility models in the financial markets. Most of the references pertain to the American stock market however some South African examples will also be supplied.

6.2 The Characteristics of Volatility

As mentioned previously variance has an important role to play in financial applications. Note however, that the variance of an asset return is not directly observable and thus one has to use statistical models in order to estimate daily, monthly as well as annual volatility. There are however a few characteristics associated with volatility that is commonly seen in asset returns. They are as follows:

1. There exists volatility clusters. i.e. volatility may be high for certain periods and low during other periods. Mandelbrot (1963) was the first to document this and stated that "... *large changes tend to be followed by large changes-of either sign- and small changes by small changes...*".
2. Volatility evolves over time in a continuous manner (Bollerslev (1986))-that is, volatility jumps are rare. This can be seen by investigating the implied volatility of option pricing models. The Black Scholes model can be used in order to determine the price of options and futures. The model requires a number of estimates of which annual standard deviation is one of them. Implied volatility is the value of the annual standard deviation that solves the Black Scholes formula when the standard deviation is the only unknown variable in the model. It can be used as an estimate of what the "market" believes the volatility of a share is.
3. Volatility behaves asymmetrically to different types of news. Black (1976) states that volatility tends to rise in response to bad news and similarly fall due to good news. Therefore it seems to react differently to a big price increase or a big price drop.

4. Volatility does not diverge to infinity- this means that there exists some fixed interval or range in which volatility varies.

The aforementioned 4 properties of volatility play a vital role in the development of volatility models. Some models were developed to specifically capture certain features mentioned before. A good example of this was the Exponential GARCH model which was developed with the purpose of capturing the asymmetry in volatility induced by large positive and negative asset returns.

6.3 Sources of ARCH and variables used to model Volatility

Volatility models were established in order to model the time varying nature of the conditional variance of time series. A number of researchers have tried to explain why conditional variances should be serially correlated. Diebold and Nerlove (1989) believe that share returns are heteroskedastic due to the existence of a serially correlated news process. Gallant, Hsieh and Tauchen (1989) offer evidence in favour of the above hypothesis however Engle, Ito and Lin (1990a,b) were unable to provide any satisfactory explanation for the dependence in the underlying news arrival process.

Lamoureux and Lastrapes (1990b) argue that heteroskedasticity found in share returns are due to the clustering in trading volumes. When they included trading volumes into their variance equation they found that the lagged residuals were not significant thus substantiating their claim. Karpoff (1987) however show that trading volumes and the price

of a share are highly correlated. This could be the reason why Lamoureux and Lastrapes (1990b) found a significant loading on the trading volume in their variance equation.

In the quest to model the conditional variance of a shares return series, researchers have unearthed numerous economic variables that are related to the conditional variance of share returns. Gallant, Rossi and Tauchen (1992) found that lagged volume were positively related to the conditional volatility on the NYSE where as Campbell (1987) and Glosten, Jagannathan and Runkle (1991) found that nominal interest rates and volatility were related. Attanasio (1991) and Attanasio and Wadhwani (1989) show that dividend yields were a significant determinant of volatility where as Engle and Rodrigues (1989) found a positive relationship between M1 money supply, the oil price and conditional volatility.

We conclude by stating a general type Regression Time Series Model and separately model the error terms as *ARCH*–*GARCH* Models. Any Regression Time Series Model can be generalized to a model of the kind

$$r_t = f(r_{t-1}, r_{t-2}, \dots, r_{t-p}; e_{t-1}, \dots, e_{t-q}; \beta_f; \mathbf{x}, \beta_x) + e_t \quad (1)$$

$$e_t = g_t \varepsilon_t$$

$$g_t = g(r_{t-1}, \dots, r_{t-u}; e_{t-1}, \dots, e_{t-v}; g_{t-1}, \dots, g_{t-w}; \beta_g)$$

where r_t is the simple or log (or excess) returns of the time series, \mathbf{x} is a set of k explanatory variables with coefficient vector β_x , $f(\cdot)$ and $g(\cdot)$ are two known functions with finite dimensional parameter vectors β_f and β_g , respectively, p, q, u, v and w are nonnegative integers; and $\{\varepsilon_t\}$ is a sequence of independent and identically distributed random variables with mean zero and variance 1. The function $g(\cdot)$ is assumed to be positive; it

governs the evolution of the volatility (i.e. conditional variance) of the innovation series e_t . For simplicity we focus on the case that the ε_t 's are standard normal or t -distributions, that is, conditional normal or t -distributions. However, it is easily seen that ε_t can be any continuous random variables with a well defined density function.

Some examples are:

1. (a) If $g(.) = \beta_1$, which is a positive constant, $k = 0$ and

$$f(.) = \sum_{i=1}^p \phi_i r_{t-i} + \sum_{i=1}^q \theta_i e_{t-i} \quad (2)$$

then model (1) reduces to the well-known *ARMA* models of Box et al (1994).

- (b) If $g(.) = \beta_1$, a constant and

$$f(.) = \sum_{i=1}^p \phi_i r_{t-i} + \sum_{i=1}^q \theta_i e_{t-i} + \mathbf{x}'\beta_x$$

then model (1) reduces to the well-known Multiple Regression *ARMA* models used in many Econometric studies.

2. If $f(.) = 0$ and

$$g^2(.) = \gamma_0 + \sum_{i=1}^v \gamma_i e_{t-i}^2, \quad (3)$$

where $\gamma_0 > 0$ and $\gamma_i \geq 0$,

then the model (1) becomes the well-known conditional autoregressive heteroskedastic (*ARCH*) model of

Engle(1982). The *ARCH* models are widely used in Finance to model the volatility of a security return.

3. If $f(.) = 0$ and

$$g^2(.) = \gamma_0 + \sum_{i=1}^v \gamma_i e_{t-i}^2 + \sum_{i=1}^w \lambda_i g_{t-i}^2, \quad (4)$$

where $\gamma_0 > 0$, $\gamma_i \geq 0$, and $\lambda_i \geq 0$,

then we have the generalized *ARCH – GARCH* model of Bollerslev(1986).

4. If $f(.) = 0$ and

$$g(.) = \exp(\gamma_0 + \sum_{i=1}^u \beta_i r_{t-i} + \sum_{j=1}^v \gamma_j e_{t-j}), \quad (5)$$

where $\gamma_0 > 0$ and $\gamma_i \geq 0$,

then model (1) becomes the stochastic volatility model in which the conditional variance of the series is related to past observations and past innovations. This model is similar to that in Tsay(1987) and can be extended to include models that allow for asymmetric responses to positive and negative innovations.

5. If $g(.) = \beta_1 > 0$, a positive constant, and

$$f(.) = \sum_{i=1}^p \phi_i r_{t-i} - \sum_{i=1}^q \theta_i e_{t-i} + \sum_{i=1}^p \sum_{j=1}^q \beta_{ij} r_{t-i} e_{t-j}, \quad (6)$$

then model (1) becomes the bilinear model of Granger and Andersen (1978) and Subba Rao (1981).

6. If

$$f(.) = \phi_0^{(i)} + \sum_{i=1}^p \phi_i^{(i)} r_{t-i} \text{ and} \quad (7)$$

$$g(.) = \sigma^{(i)} > 0, \text{ for } x_{i-1} \leq r_{t-d} \leq x_i,$$

where d is a positive integer and x_i 's are real numbers satisfying

$$-\infty = x_0 < x_1 < \dots < x_k = \infty,$$

then the model (1) becomes the threshold autoregressive TAR model of Tong (1987,1990).

7. Model (1) also provides the framework to combine different time series models, for example

$$\begin{aligned} f(.) &= \phi_0 + \phi_1 r_{t-1} - \theta_1 e_{t-1} \text{ and} \\ g(.) &= \omega_0 + \omega_1 e_{t-1} \end{aligned} \tag{8}$$

so that r_t is an $ARMA$ process with a concurrent bilinear innovation. The model (1) may be vector valued processes leading to $VARMA$ processes. This is of particular interest in portfolio analysis which will incorporate the Troskie Hossain formulations for the covariance error structure matrix Ξ (i.e. the Improved Sharpe).

6.4 Structure of A Model

We have established that r_t is the log return of an asset at time index t . The purpose to volatility modelling is that the series $\{r_t\}$ is not independent and thus has some sort of dependence structure. Thus we are suggesting that the returns series are serially uncorrelated but dependent. Hence the volatility models endeavour to encapsulate this dependence feature in the return series. To portray a clear perspective of volatility models, it is informative to consider the conditional mean and conditional variance of r_t given F_{t-1} that is

$$\mu_t = E(r_t | F_{t-1}), \quad \sigma_t^2 = Var(r_t | F_{t-1}) = E[(r_t - \mu_t)^2 | F_{t-1}] \tag{9}$$

where F_{t-1} denotes the information set available at time $t - 1$. Typically F_{t-1} consists of all linear functions of the past returns. The equation for μ_t in (9) should be simple since the serial dependence of stock returns is weak if it exists at all. Thus it is safe to assume that r_t follows a simple time series model -, an $ARMA(p, q)$ model:

$$r_t = \mu_t + a_t$$

$$\text{where } \mu_t = \phi_0 + \sum_{i=1}^p \phi_i r_{t-i} - \sum_{i=1}^q \theta_i a_{t-i} \quad (10)$$

Combining equations (9) and (10) we have

$$\sigma_t^2 = Var(r_t | F_{t-1}) = Var(a_t | F_{t-1}) \quad (11)$$

where a_t is the shock or mean-corrected return of an asset return at time t . Equation (11) is referred to as the volatility equation for r_t and equation (10) is the mean equation.

Conditional heteroskedastic models can be classified into 2 categories:

1. The 1st category are the GARCH models: these models use an exact function to govern the evolution of σ_t^2
2. The 2nd category are the stochastic volatility models: These models use a stochastic equation to describe σ_t^2

When introducing volatility models, it is assumed that the conditional mean is given, whilst in empirical studies the conditional mean and variance are estimated.

For model simplification we formally introduce the volatility models as

$$e_t = r_t, \quad (9)$$

that is, we model only the residuals, assuming that the model in the background may be any form of the model given by equation (1). Thus our model is

$$r_t = \sqrt{h_t}\varepsilon_t \quad \text{where} \quad h_t = \text{Var}(r_t|F_{t-1}), \quad (10)$$

where ε_t is assumed to be a **white noise process with variance 1** and F_{t-1} is the information set available at time $t - 1$. Often ε_t is modelled as a **standard normal** random variate or as a **standardised Students t distribution**.

6.5 Different Specifications of the error distribution

A number of distributions have been suggested in order to model the error process $\{\varepsilon_t\}$ of a volatility model. Bollerslev (1987) suggests using a standardized Student t distribution with v degrees of freedom, where v has to be estimated from the data. Other parametric densities have also been considered. Jorion (1988) suggested using a normal-Poisson mixture distribution, Baillie and Bollerslev (1989a) proposed using the power exponential distribution, Hsieh (1989a) used the normal-lognormal mixture distribution, Nelson (1990a) used a generalised exponential distribution and McCulloch (1985) used an infinite variance leptokurtic stable Pareto distribution in order to model the distribution of the error process.

Semi-parametric density estimation techniques have been proposed by Gallant and Nychka (1987), Gallant and Tauchen (1989), Gallant, Hsieh and Tauchen (1989), Gallant, Rossi and Tauchen (1992) and Engle and Gonzalez-Rivera (1991). In general the semi-parametric approach leads to a loss in asymptotic efficiency when one compares the estimation approach to maximum likelihood estimation with a correctly specified density function for the error process of the volatility model.

6.5.1 ARCH Models with a t distribution

Bollerslev (1987), Hsieh (1989a), Baillie and Bollerslev (1989) and Palm and Vlaar (1997) show that the Student's t specification for the innovations better captures the observed kurtosis in returns data. Milhoj (1985) and Bollerslev (1986) however state that the specification remains inadequate for many financial time series since the observed time series have fatter tails than the Student's t distribution.

6.5.2 ARCH Models with a Skewed Standardised t distribution

Bollerslev (1987), Hsieh (1989), Ballie and Bollerslev (1989) and Palm and Vlaar (1997) show that the Students t specification for the error distribution better captures the observed kurtosis in share returns data. Share returns are often skewed. Symmetric distributions such as the normal and Students t distribution do not adequately capture this property of share returns. Fernández and Steel (1998) proposed using a skewed Students t distribution to model the distribution of share returns. Lambert and Laurent (2001) used this idea in order to model the error distribution of a volatility model as a standardised skewed Students t distribution.

A Skewed density function of a symmetric unimodal density function $g(u)$ (where U is continuous) can be constructed by transforming U to $\epsilon = X\xi|U| - (1 - X)\frac{1}{\xi}|U|$ where X is a Bernoulli random variable, with probability of success $\frac{\xi^2}{1+\xi^2}$. We can show that the unconditional density $f(\epsilon|\xi)$ of ϵ is equal to:

$$\begin{aligned} f(\epsilon|\xi) &= \Pr(X = 0)g(\epsilon|\xi, x = 0) + \Pr(X = 1)g(\epsilon|\xi, x = 1) \\ &= \frac{2}{\xi + \frac{1}{\xi}} \left[g(\epsilon\xi) I_{(-\infty, 0)}(\epsilon) + g\left(\frac{\epsilon}{\xi}\right) I_{[0, \infty)}(\epsilon) \right] \end{aligned} \quad (1)$$

where $I(\epsilon)$ is an indicator variable such that $I_{(a,b)}(\epsilon) = 1$ if $a < \epsilon < b$. Fernández and Steel (1998) show that if the r^{th} moment of $g(\cdot)$ exists then the moments of $f(\epsilon|\xi)$ are defined and is equal to:

$$E(\epsilon^r|\xi) = M_r \frac{\xi^{r+1} + \frac{(-1)^r}{\xi^{r+1}}}{\xi + \frac{1}{\xi}} \quad \text{where} \quad M_r = \int_0^\infty 2s^r g(s) ds \quad (2)$$

ϵ does not have a zero mean or a variance of 1. Lambert and Laurent (2001) proposed standardising ϵ so that it could be used as the error distribution of a volatility model

(Lambert and Laurent (2001) used a $GARCH(1, 1)$ to model the conditional volatility. GARCH models will be covered in the following section.) If z_t is the standardised random variable then the density function of z_t is:

$$\left(\frac{2s}{\xi + \frac{1}{\xi}} \right) \left[g \left[\xi (sz + m) \right] I_{(-\infty, 0)} \left(z + \frac{m}{s} \right) + g \left[\frac{(sz + m)}{\xi} \right] I_{[0, \infty)} \left(z + \frac{m}{s} \right) \right] \quad (3)$$

where $z_t = \frac{\epsilon_t - m}{s}$, $m = E(\epsilon_t | \xi)$ and $s^2 = Var(\epsilon_t | \xi)$. If U is assumed to be a standardised Students t distribution then the contribution of the t^{th} observation to the log likelihood function is equal to:

$$\ln \left(\frac{\Gamma(\frac{v+1}{2})}{\Gamma(\frac{v}{2}) \sqrt{(v-2)\pi}} \right) + \ln \left(\frac{2\xi s}{1 + \xi^2} \right) - \frac{1}{2} \left\{ \ln(\sigma_t^2) + (1+v) \left(1 + \frac{(sz_t + m)^2}{v-2} \xi^{-2A_t} \right) \right\} \quad (4)$$

where $z_t = \frac{\epsilon_t - m}{s}$, $m = \sqrt{\frac{v-2}{\pi}} \frac{\Gamma(\frac{v-1}{2})}{\Gamma(\frac{v}{2})} \left(\xi - \frac{1}{\xi} \right)$, $s^2 = \left(\xi^2 + \frac{1}{\xi^2} - 1 \right) - m^2$ and

$$A_t = \begin{cases} 1 & \text{if } z_t \geq \frac{m}{s} \\ -1 & \text{if } z_t < \frac{m}{s} \end{cases}$$

Lambert and Laurent (2001) undertook a Monte Carlo simulation in order to test the efficiency of the MLE estimation procedure for the $GARCH(1, 1)$ model. They found that the parameter estimates showed small biases when one wrongly assumes that the error distribution is a skewed t distribution. It was however noted that the specification of the skewed t distribution as the error distribution was superior to Quasi Maximum Likelihood estimation (i.e. the normality assumption) since it allows for both skewness and kurtosis to be incorporated into the model. The benefits of the skewed Standardised t distribution was demonstrated by examining twelve years of **daily** returns data on the NASDAQ from January 1985 to December 1996. Pearson goodness of fit tests were used to test assumptions about the error distribution of the GARCH process. It was found that both the normal

and the Students t distribution could not adequately model the returns series. The skewed Standardised t distribution however was found to be adequate.

6.6 Linear Regression with ARCH error terms

Suppose that one wishes to estimate the parameters of a regression model with *ARCH* errors. Let the regression equation be:

$$Y_t = X_t' \beta + e_t = X_t' \beta + \sqrt{h_t} \varepsilon_t \quad (1)$$

where $e_t = \sqrt{h_t} \varepsilon_t$, X_t contains k exogenous variables and $h_t = \alpha_0 + \alpha_1(Y_{t-1} - X_{t-1}' \beta)^2 + \alpha_2(Y_{t-2} - X_{t-2}' \beta)^2 + \dots + \alpha_p(Y_{t-p} - X_{t-p}' \beta)^2$ and $\{\varepsilon_t\}$ is a i.i.d. random variable with mean 0 and variance 1. If $\varepsilon_t \sim N(0, 1)$ independent of X_t then the conditional distribution $Y_t|X_t \sim N(X_t' \beta, h_t)$ and thus the conditional density is:

$$f_{Y_t|X_t}(y_t|x_t) = \frac{1}{\sqrt{2\pi h_t}} \exp \left(-\frac{(y_t - x_t' \beta)^2}{2h_t} \right) \quad (2)$$

The conditional log likelihood is:

$$l_c(y_{p+1}, y_{p+2}, \dots, y_T | \alpha, \beta) = \sum_{t=p+1}^T \left\{ -\frac{1}{2} \log(2\pi) - \frac{1}{2} \log(h_t) - \frac{1}{2} \frac{(y_t - x_t' \beta)^2}{h_t} \right\} \quad (3)$$

This conditional log likelihood can be evaluated by using the Berndt, Hall and Hausman (1974) algorithm. The optimisation algorithm uses numerical differentiation in order to solve for the unknown parameters.

Weiss (1986), Bollerslev and Wooldridge (1992) and Glosten, Jagannathan and Runkle (1989) show that maximising a normal log likelihood will provide **consistent** estimates even if the residuals of the fitted model are not normally distributed (provided that the resid-

uals has a zero mean and unit variance.) They stress however that the standard errors of the estimates will have to be adjusted.

The above estimation procedure is known as Quasi Maximum likelihood estimation. Engle and Gloria Gonzalez Rivera (1991) investigated the relative efficiency of quasi maximum likelihood estimates compared to maximum likelihood estimation for the *GARCH*(1, 1) model by considering different gamma and t distribution representations of the residual series. They found that *QML* ensures the consistency of the parameter estimates however they stress that the loss of efficiency due to the normality assumption could be significant.

6.7 Other ARCH specifications

6.7.1 The ARCH in Mean Model

Much of financial economic theory relies on the assumption of the existence of a positive relationship between expected return and risk (measured by means of the variance of a shares returns). For example, the Capital Asset Pricing Model (CAPM) of Sharpe (1964), Lintner (1965), Mossin (1966), Merton (1973) or the Consumption Based CAPM of Breeden (1979), Lucas (1978), the Arbitrage Pricing Theory (APT) of Ross (1976), Chamberlain and Rothschild (1983). Due to this relationship researchers developed the ARCH-M or ARCH in mean model in order to model the conditional expectation of a share as a function of the conditional volatility. Formally the model can be expressed as:

$$r_t = \mu_t + \delta h_t + \varepsilon_t \quad (4)$$

where δ is a risk aversion factor and all of the common assumptions hold as in the above model specification. Intuitively δ should be non-negative however as shown by Baillie and DeGennaro (1990), the sign of δ is dependent on the assumption of the error distribution of the volatility model. They found that the δ coefficient changes from being significant at the 5% level to being insignificant by simply changing the error distribution assumption from being normal to being Students t. Bollerslev and Woodridge (1991) found similar results by using robust standard errors. Further evidence against the ARCH-M specification was provided by Glosten, Jagannathan and Runkle (1991). They found that the sign of δ is sensitive to the specification of the mean and variance equation. As a final criticism, it is important to note the findings of Pagan and Ullah (1998). They state that the parameters in the conditional mean equation (of the ARCH-M) are not asymptotically independent of the variance equation and thus any misspecification in the variance equation leads to inconsistent and biased estimates of the parameters in the mean equation. In many cases one should be careful when using the ARCH-M model due to the many problems associated with consistency and the sign of the δ coefficient.

6.8 The weaknesses of the ARCH Model

The following are the weaknesses of the ARCH methodology:

1. The model is sign independent with respect to returns at time t and thus does not allow one to model assymetric information flows. (Black (1976)). The model assumes that

positive and negative shocks have the same effects on volatility because it depends on the square of the previous shocks.

2. Due to the stationarity assumptions many of the parameters are restrictively constrained. This could cause convergence problems when estimating the parameter values by means of *MLE*. For example, α_1^2 of an *ARCH*(1) model must be in the interval $[0, \frac{1}{3}]$ if the series is to have a finite 4th moment. The constraint becomes complicated for higher order ARCH models
3. ARCH models often over-predict volatility, since they respond slowly to outlier returns.
4. ARCH models are mechanical in nature and does not provide any new financial insight to portfolio and risk management. They only provide a way to describe the behaviour of the conditional variance. No indication is given as to what causes such behaviour to occur.

6.9 The GARCH Model

Bollerslev (1986) developed the Generalized ARCH models since it was found that often ARCH models required relatively long lag structures in the conditional variance equation. The GARCH model is an extension of the univariate ARCH similar to the way in which ARMA models are extensions of AR and MA models. The model can be defined as follows:

Let a_t be the mean adjusted return of a share at time t and r_t is the return of a share at time t such that $a_t = r_t - \mu_t$. It is then said that a_t follows a $GARCH(p, q)$ process if and only if:

$$a_t = \sigma_t \varepsilon_t \quad \text{with} \quad \sigma_t^2 = \alpha_0 + \sum_{i=1}^p \alpha_i a_{t-i}^2 + \sum_{j=1}^q \beta_j \sigma_{t-j}^2 \quad (1)$$

where ε_t is a white noise random variable such that $E(\varepsilon_t) = 0$ and $Var(\varepsilon_t) = 1$, $\alpha_0 > 0$, $\alpha_i \geq 0$, $\beta_j \geq 0$, and $\sum_{i=1}^{\max(p,q)} (\alpha_i + \beta_i) < 1$. Here it is understood that $\alpha_i \geq 0$ for $i > p$ and $\beta_j = 0$ for $j > q$. The latter constraint on $\alpha_i + \beta_i$ implies that the unconditional variance of a_t is finite, whereas its conditional variance σ_t^2 evolves over time. Notice that in this formulation p represents the number of ARCH terms where as q represents the number of GARCH terms. Notice that equation (1) reduces to a pure $ARCH(p)$ term when $q = 0$.

If a_t is a $GARCH(p, q)$ process it can be shown that a_t^2 is an $ARMA(\max(p, q), q)$ process. Box Jenkins methods and the *ESACF* methodology could be used in order to identify the order of the GARCH process. This can be seen by rearranging equation (1) as follows:

$$\sigma_t^2 = \alpha_0 + (\alpha_1 a_{t-1}^2 + \dots + \alpha_p a_{t-p}^2) + (\beta_1 \sigma_{t-1}^2 + \dots + \beta_q \sigma_{t-q}^2) \quad (2)$$

Now adding a_t^2 to both sides and rearranging one gets:

$$\begin{aligned} \sigma_t^2 + a_t^2 &= \alpha_0 + (\alpha_1 a_{t-1}^2 + \dots + \alpha_p a_{t-p}^2) + (\beta_1 \sigma_{t-1}^2 + \dots + \beta_q \sigma_{t-q}^2) + a_t^2 \\ &= \alpha_0 - \beta_1 (a_{t-1}^2 - \sigma_{t-1}^2) - \beta_2 (a_{t-2}^2 - \sigma_{t-2}^2) - \dots - \beta_q (a_{t-q}^2 - \sigma_{t-q}^2) + \\ &\quad \beta_1 a_{t-1}^2 + \beta_2 a_{t-2}^2 + \dots + \beta_q a_{t-q}^2 + (\alpha_1 a_{t-1}^2 + \alpha_2 a_{t-2}^2 + \dots + \alpha_p a_{t-p}^2) + a_t^2 \end{aligned} \quad (3)$$

Now rearranging:

$$a_t^2 = \alpha_0 + (a_t^2 - \sigma_t^2) - \beta_1(a_{t-1}^2 - \sigma_{t-1}^2) - \beta_2(a_{t-2}^2 - \sigma_{t-2}^2) - \dots - \beta_q(a_{t-q}^2 - \sigma_{t-q}^2) + \\ + \beta_1 a_{t-1}^2 + \beta_2 a_{t-2}^2 + \dots + \beta_q a_{t-q}^2 + \alpha_1 a_{t-1}^2 + \alpha_2 a_{t-2}^2 + \dots + \alpha_p a_{t-p}^2$$

Such that:

$$a_t^2 = \alpha_0 + ((\alpha_1 + \beta_1)a_{t-1}^2 + \dots + (\alpha_p + \beta_p)a_{t-p}^2) + (w_t - \beta_1 w_{t-1} - \dots - \beta_q w_{t-q}) \quad (4)$$

where $w_t = a_{t-1}^2 - \sigma_{t-1}^2$ so that

$$a_t^2 = \alpha_0 + \sum_{i=1}^{\max(p,q)} (\alpha_i + \beta_i) a_{t-i}^2 + w_t - \sum_{j=1}^q \beta_j w_{t-j} \quad (5)$$

From this one can see that if a_t is a $GARCH(p, q)$ process then a_t^2 is an $ARMA(a, b)$ process where the AR coefficients are $\alpha_i + \beta_i$, the MA coefficients are β_i , $a = \max(p, q)$ and $b = q$. The $GARCH$ process is thus only stationary if a_t^2 is stationary implying that $|\alpha_i + \beta_i| < 1$ and $|\beta_i| < 1$. Since $h_t \geq 0$, the stationarity conditions are: $0 < \alpha_i + \beta_i < 1$ and $0 < \beta_i < 1$ and thus $0 < \sum \alpha_i + \sum \beta_i < 1$. Using the unconditional mean of an ARMA model, we have

$$E(a_t^2) = \frac{\alpha_0}{1 - \sum_{i=1}^{\max(p,q)} (\alpha_i + \beta_i)}$$

provided that the denominator of the prior fraction is positive.

6.10 The GARCH(1,1) Model

The model is specified as follows:

$$a_t = \sigma_t \varepsilon_t \quad \text{where} \quad \sigma_t^2 = \alpha_0 + \alpha_1 a_{t-1}^2 + \beta_1 \sigma_{t-1}^2 \quad (1)$$

or

$$a_t^2 = \alpha_0 + (\alpha_1 + \beta_1) a_{t-1}^2 + w_t - \beta_1 w_{t-1} \quad (2)$$

Assuming that the error distribution is a $N(0, 1)$ random variable then the log likelihood function is identical to equation 2 in the "Estimation" section, although the summation starts at 2 and not $p + 1$. Note however that since σ_t^2 depends on the first lagged conditional variance one requires an estimate of σ_1^2 . The unconditional sample estimate is often used as a simple approximation (as recommended by Bollerslev (1986)). σ_1^2 could also be set equal to zero.

6.11 Limitations of the GARCH Model

GARCH model are generalizations of ARCH models and thus have many of the same limitations as the ARCH model. They are however able to model the conditional variance of a time series by using relatively few parameters. A $GARCH(1, 1)$ is most often used since in general it is difficult to specify the values of p and q . Notice also that the GARCH model treats both positive and negative returns in the same way since the square of lagged returns are used in order to model the conditional variance equation of shares returns.

Assymetric volatility models were developed in order to incorporate the sign of lagged returns into the specification of the variance equation. Threshold ARCH (TARCH) (Zakoin (1990), and (Glosten, Jaganath, and Runkle (1993)) and Exponential $GARCH$

(*EGARCH*) are two types of asymmetric volatility models. The following section briefly discusses the *EGARCH* methodology.

6.12 Empirical Investigation: Single Index

6.12.1 The data

For this empirical analysis, the data used was the SIDS batch.

6.12.2 Study Objectives

Up to now, we have established time series, regression and state space models for R_1, R_2, R_3, R_4, R_5 and R_9 . Our next objective is to produce volatility *GARCH*(1, 1) models for each of these nine response variables. Furthermore a comparative study will be conducted in order to establish which modelling approach, state space or *GARCH* modelling, fits the data 'best', which leads us to which modelling approach is superior. This objective will be further explored in the portfolio theory chapter to follow with regards to mean-variance frontiers For now we only examine and compare the two type of modelling paradigms for each response.

6.12.3 Methodology

There are 3 steps when building an ARCH Model:

1. First one needs to establish an econometric model for the return series in order to eliminate any linear dependence in the data , e.g. a time series or regression model -

$ARMA(1, 1)$ for example. Then use the residual series of the model to test for ARCH effects.

2. Then one needs to specify the ARCH order and then execute the estimation process
3. Finally an empirical and significance check of the ARCH model fit needs to be reviewed carefully and then to refine it if necessary.

From chapter 4, we showed that the response variables R_1, R_3, R_4, R_8 and R_9 had AR time series components, while the rest were in the form of simple regression models accompanied by only one explanatory, JSE. These nine models found under the time series context will now be taken in their respective forms and refitted with their explanatory and AR components under the $GARCH(1, 1)$ environ.

Once fitting the respective models for all nine responses, the model selection procedure is based on the R^2 , Adjusted R^2 and the Durbin Watson Statistic. Once the appropriate $GARCH$ models are fit for the respective responses, the most efficient model is opted for based on the aforementioned selection criteria. Subsequent to the selection procedure we will conclude with a comparison between State Space and $GARCH$ modelling in terms of model fit.

6.12.4 Primary Findings

From the previous empirical study state space models were established for the nine financial stocks in our portfolio. The above outlined methodology was applied for each of the time series models acquired from chapter 4. We highlight the results from applying the

$GARCH(1,1)$ estimation procedures for our portfolio. Table 6.1 below illustrates a summary of the model selection criteria from fitting the new models

GARCH(1,1) Models	Model Selection Criteria						
	Adj-R ²	S.E. of reg	D-W stat	AIC	BIC	F-stat	Prob(F-stat)
Model for R1 (Model 1)	0.6824	0.0563	1.9696	-3.0350	-2.8850	53.6338	0.0000
Model for R2 (Model 2)	0.1807	0.1241	2.1609	-1.3453	-1.2461	9.7321	0.0000
Model for R3 (Model 3)	0.1959	0.0916	1.9697	-2.0189	-1.9026	8.9999	0.0000
Model for R4 (Model 4)	0.4534	0.0599	2.0272	-2.7792	-2.6463	22.5159	0.0000
Model for R5 (Model 5)	0.0714	0.0820	1.9845	-2.2592	-2.1433	2.4590	0.0258
Model for R6 (Model 6)	0.3119	0.0919	1.8860	-2.0600	-1.9607	18.9462	0.0000
Model for R7 (Model 7)	0.3559	0.0792	1.8817	-2.2421	-2.1593	28.3461	0.0000
Model for R8 (Model 8)	0.2825	0.0659	2.1359	-2.5777	-2.4763	16.1215	0.0000
Model for R9 (Model 9)	0.2193	0.0811	2.0367	-2.2172	-2.1172	12.0086	0.0000

Table 6.1: Summary statistics for Garch(1,1) models

Before any further analysis of the results, table 6.2 below presents the $GARCH(1,1)$ forms adopted for each of the responses variables R_1 through to R_9 .

GARCH(1,1) Models	Definition Of GARCH(1,1) Structure
Model for R1 (Model 1)	$R1 = C(1)*\sqrt{GARCH} + C(2) + C(3)*JSE + [AR(2)=C(4)]$ $GARCH = C(5) + C(6)*RESID(-1)^2 + C(7)*GARCH(-1) + C(8)*JSE$
Model for R2 (Model 2)	$R2 = C(1)*\sqrt{GARCH} + C(2) + C(3)*JSE$ $GARCH = C(4) + C(5)*RESID(-1)^2 + C(6)*GARCH(-1)$
Model for R3 (Model 3)	$R3 = C(1) + C(2)*JSE + [AR(1)=C(3)]$ $GARCH = C(4) + C(5)*RESID(-1)^2 + C(6)*GARCH(-1)$
Model for R4 (Model 4)	$R4 = C(1)*\sqrt{GARCH} + C(2) + C(3)*JSE + [AR(1)=C(4)]$ $GARCH = C(5) + C(6)*RESID(-1)^2 + C(7)*GARCH(-1)$
Model for R5 (Model 5)	$R5 = C(1)*\log(GARCH) + C(2) + C(3)*JSE$ $GARCH = C(4) + C(5)*RESID(-1)^2 + C(6)*GARCH(-1)$
Model for R6 (Model 6)	$R6 = C(1) + C(2)*JSE$ $GARCH = C(3) + C(4)*RESID(-1)^2 + C(5)*GARCH(-1)$
Model for R7 (Model 7)	$R7 = C(1) + C(2)*JSE$ $GARCH = C(3) + C(4)*RESID(-1)^2 + C(5)*GARCH(-1)$
Model for R8 (Model 8)	$R8 = C(1) + C(2)*JSE + [AR(6)=C(3)]$ $GARCH = C(4) + C(5)*RESID(-1)^2 + C(6)*GARCH(-1)$
Model for R9 (Model 9)	$R9 = C(1) + C(2)*JSE + [AR(2)=C(3)]$ $GARCH = C(4) + C(5)*RESID(-1)^2 + C(6)*GARCH(-1)$

Table 6.2: Garch(1,1) structure of responses

Examining the results from table 6.1 above, it is blatantly clear that only 2 out of 9 $Adj - R^2$ statistics, namely variables R_4 and R_5 , increased, thus explaining more variation than their state space counterparts. Accompanying the higher achieved $Adj - R^2$, these $GARCH(1,1)$

models have also successfully reduced the SE of Regression for R_4 and R_5 . Meanwhile for the remaining 7 responses, the state space models have maintained their superiority in terms of $Adj - R^2$ and the SE of Regression. Thus the state space models provide better explanatory power in terms of higher R^2 and lower error variances of regression. Next we examine the significance and stability of the new updated coefficients of the JSE in table 6.3 below, since these are the main coefficients to be used as inputs for the construction of our mean-variance frontier for $GARCH(1, 1)$.

GARCH(1,1) Models	Coefficient of JSE			
	Coefficient (β_i)	Std. Error	t-Statistic	Prob.
Model for R1 (Model 1)	1.4735	0.0514	28.6926	0.0000
Model for R2 (Model 2)	1.1383	0.1210	9.4059	0.0000
Model for R3 (Model 3)	0.6320	0.0962	6.5726	0.0000
Model for R4 (Model 4)	0.8285	0.0724	11.4401	0.0000
Model for R5 (Model 5)	0.2647	0.0772	3.4299	0.0006
Model for R6 (Model 6)	1.0014	0.0881	11.3640	0.0000
Model for R7 (Model 7)	0.9803	0.1035	9.4739	0.0000
Model for R8 (Model 8)	0.7424	0.0726	10.2266	0.0000
Model for R9 (Model 9)	0.7477	0.1088	6.8750	0.0000

Table 6.3: JSE Coefficients from GARCH(1,1) models

It is clear from table 6.3 above that the coefficients are all significant at the 5% level of significance. Thus the $GARCH(1, 1)$ models are all significant. The beta values computed under $GARCH(1, 1)$ are generally more stable and have larger t-stats and lower std. errors compared to that of the state space models for R_1, R_2, R_4, R_5, R_8 and R_9 . Table 6.4 below compares this finding below. What remains to be seen is, which modelling approach is superior with regards to the risk-return frontier. The state space models have demonstrated superior variation explanation whilst the $GARCH(1, 1)$ models have demonstrated superior beta estimates and t-statistics.

Coefficient of JSE				
GARCH(1,1) Models	Coefficient (β_i)	Std. Error	t-Statistic	Prob.
Model for R1 (Model 1)	1.4735	0.0514	28.6926	0.0000
Model for R2 (Model 2)	1.1383	0.1210	9.4059	0.0000
Model for R3 (Model 3)	0.6320	0.0962	6.5726	0.0000
Model for R4 (Model 4)	0.8285	0.0724	11.4401	0.0000
Model for R5 (Model 5)	0.2647	0.0772	3.4299	0.0006
Model for R6 (Model 6)	1.0014	0.0881	11.3640	0.0000
Model for R7 (Model 7)	0.9803	0.1035	9.4739	0.0000
Model for R8 (Model 8)	0.7424	0.0726	10.2266	0.0000
Model for R9 (Model 9)	0.7477	0.1088	6.8750	0.0000
ARMA State Space Model Coefficients				
Coefficient of JSE C(2)				
ARMA State Space Models	C(2)	Std. Error	t-Statistic	Prob.
Model for R1 (Model 1)	1.4199	0.0346	41.0482	0.0000
Model for R3 (Model 3)	0.7044	0.1169	6.0248	0.0000
Model for R4 (Model 4)	0.8199	0.0821	9.9883	0.0000
Model for R8 (Model 8)	0.7340	0.0740	9.9242	0.0000
Model for R9 (Model 9)	0.7460	0.1059	7.0460	0.0000
Recursive Coefficient State Space Model Coefficients				
Coefficient of JSE SV2				
Recursive Coefficient Models	Final SV2	Std. Error	t-Statistic	Prob.
Model for R2 (Model 2)	1.0359	0.1478	7.0083	0.0000
Model for R5 (Model 5)	0.3115	0.0985	3.1606	0.0018
Model for R6 (Model 6)	1.0776	0.1092	9.8721	0.0000
Model for R7 (Model 7)	1.0141	0.0944	10.7397	0.0000

Table 6.4: Coefficients for JSE - GARCH vs State Space

Thus it remains to be seen whether or not in fact state space methodology is superior to $GARCH(1, 1)$ in a portfolio optimisation sense. We can only determine this once efficient frontiers have been constructed. The conclusions follow regarding the above findings in the subsequent section.

6.12.5 Conclusions

Having completed implementing the $GARCH(1, 1)$ methodologies on the respective responses, the following is concluded: that significant $GARCH(1, 1)$ models have been found for all nine stocks of our portfolio. Furthermore the $GARCH(1, 1)$ models acquired appear to provide better explanatory power in terms of high R^2 values and low error variances of the models than the time series models for the stocks R_4 and R_5 .. Furthermore

we conclude that State Space models are more superior with regards to a modelling sense. But it was the $GARCH(1, 1)$ models that demonstrated superior beta estimates and values compared to state space models. Hence it still remains to be seen which is superior with regards to the mean-variance frontier. The betas and regression SE's computed for the 9 stocks will be used as inputs for the computation of the efficient frontier for the single index state space model. The variance JSE and the mean log returns of the nine stocks will also be required as inputs for the computation of the $GARCH(1, 1)$ method's mean-variance frontier.

6.13 Empirical Investigation: Multiple Index

6.13.1 The data

For this empirical analysis, the data used was the MIDS batch.

6.13.2 Study Objectives

We have established (MI) time series, (MI) regression and (MI) state space models for $R_1, R_2, R_3, R_4, R_5, R_6, R_7, R_8$ and R_9 . The next step is to produce (MI) $GARCH(1,1)$ models for each of these nine response variables. Furthermore a comparative study will be conducted in order to establish which modelling approach, state space or GARCH modelling, fits the data 'best'. This objective will be further explored in the portfolio theory chapter to follow with regards to mean-variance frontiers where it will be finally estab-

lished as to which modelling approach is superior in a portfolio optimisation environment. For now we only examine and compare the modelling paradigms for each response.

6.13.3 Methodology

Recall from the previous methodology section that there are 3 steps when building an ARCH Model:

1. First one needs to establish an econometric model for the return series in order to eliminate any linear dependence in the data, e.g. a time series or regression model - $ARMA(1, 1)$ for example. Then use the residual series of the model to test for ARCH effects.
2. Then one needs to specify the ARCH order and then execute the estimation process.
3. Finally an empirical and significance check of the ARCH model fit needs to be reviewed carefully and then to refine it if necessary.

In chapter 4, it was shown that the response variables R_1, R_3, R_4, R_8 and R_9 had AR time series components, while the rest were in the form of simple (MI) regression models. These nine models found under the time series context will now be taken in their respective forms and refitted with their explanatory and AR components under the (MI) $GARCH(1, 1)$ setting.

Once fitting the respective models for all nine responses, the model selection procedure is based on the R^2 , Adjusted R^2 and the Durbin Watson Statistic. Once the appropriate $GARCH$ models are fit for the respective responses, the most efficient model is opted for

based on the aforementioned selection criteria. Subsequent to the selection procedure we will conclude with a comparison between state space and GARCH modelling in the (MI) context.

6.13.4 Primary Findings

From the previous chapter's (MI) empirical study, state space models were established for the 9 financial stocks in our portfolio. The above outlined methodology was applied for each of the time series models acquired from chapter 4. We highlight the results from applying the (MI) $GARCH(1, 1)$ estimation procedures for our portfolio. Table 6.5 below illustrates a summary of the model selection criteria from fitting the new (MI) models

GARCH(1,1) Models	Model Selection Criteria						
	Adj-R ²	S.E. of reg	D-W stat	AIC	BIC	F-stat	Prob(F-stat)
Model for R1 (Model 1)	0.7041	0.0543	2.0291	-2.9987	-2.8487	59.2843	0.0000
Model for R2 (Model 2)	0.2258	0.1206	2.1698	-1.3539	-1.2215	9.2498	0.0000
Model for R3 (Model 3)	0.2853	0.0864	1.8793	-2.0840	-1.9511	12.2323	0.0000
Model for R4 (Model 4)	0.4467	0.0592	2.0539	-2.7672	-2.6178	20.8792	0.0000
Model for R5 (Model 5)	0.0528	0.0816	2.0559	-2.1595	-2.0271	2.5770	0.0147
Model for R6 (Model 6)	0.3681	0.0880	1.9446	-2.0348	-1.9024	17.4754	0.0000
Model for R7 (Model 7)	0.3680	0.0784	1.8383	-2.2451	-2.1292	20.2125	0.0000
Model for R8 (Model 8)	0.3228	0.0640	2.0833	-2.6235	-2.4713	12.4419	0.0000
Model for R9 (Model 9)	0.2436	0.0799	2.0766	-2.2879	-2.1546	10.0177	0.0000

Table 6.5: Summary statistics for (MI) Garch(1,1) models

Before any further analysis of the results, table 6.6 below presents the $MI - GARCH(1, 1)$ forms adopted for each of the responses variables R_1 through to R_9 .

GARCH(1,1) Models	Definition Of GARCH(1,1) Structure
Model for R1 (Model 1)	$R1 = C(1) + C(2)'X5 + C(3)'X6 + C(4)'X7 + [AR(2)=C(5)]$ $GARCH = C(6) + C(7)'RESID(-1)^2 + C(8)'GARCH(-1)$
Model for R2 (Model 2)	$R2 = C(1)'@SQRT(GARCH) + C(2) + C(3)'X5 + C(4)'X6 + C(5)'X7$ $GARCH = C(6) + C(7)'RESID(-1)^2 + C(8)'GARCH(-1)$
Model for R3 (Model 3)	$R3 = C(1) + C(2)'X5 + C(3)'X6 + C(4)'X7 + [AR(1)=C(5)]$ $GARCH = C(6) + C(7)'RESID(-1)^2 + C(8)'GARCH(-1)$
Model for R4 (Model 4)	$R4 = C(1)'@SQRT(GARCH) + C(2) + C(3)'X5 + C(4)'X6 + C(5)'X7 + [AR(1)=C(6)]$ $GARCH = C(7) + C(8)'RESID(-1)^2 + C(9)'GARCH(-1)$
Model for R5 (Model 5)	$R5 = C(1)'@SQRT(GARCH) + C(2) + C(3)'X5 + C(4)'X6 + C(5)'X7$ $GARCH = C(6) + C(7)'RESID(-1)^2 + C(8)'GARCH(-1)$
Model for R6 (Model 6)	$R6 = C(1)'LOG(GARCH) + C(2) + C(3)'X5 + C(4)'X6 + C(5)'X7$ $GARCH = C(6) + C(7)'RESID(-1)^2 + C(8)'GARCH(-1)$
Model for R7 (Model 7)	$R7 = C(1) + C(2)'X5 + C(3)'X6 + C(4)'X7$ $GARCH = C(5) + C(6)'RESID(-1)^2 + C(7)'GARCH(-1)$
Model for R8 (Model 8)	$R8 = C(1)'LOG(GARCH) + C(2) + C(3)'X5 + C(4)'X6 + C(5)'X7 + [AR(6)=C(6)]$ $GARCH = C(7) + C(8)'RESID(-1)^2 + C(9)'GARCH(-1)$
Model for R9 (Model 9)	$R9 = C(1) + C(2)'X5 + C(3)'X6 + C(4)'X7 + [AR(2)=C(5)]$ $GARCH = C(6) + C(7)'RESID(-1)^2 + C(8)'GARCH(-1)$

Table 6.6: MI-Garch(1,1) structure of responses

Examining the results from table 6.5 above, it is blatantly clear that only 2 out of 9 $Adj - R^2$ statistics, namely variables R_4 and R_2 , increased, thus explaining more variation than their state space counterparts. Accompanying the higher achieved $Adj - R^2$, these $GARCH(1, 1)$ models have also successfully reduced the SE of Regression for R_4 and R_5 . Meanwhile for the remaining 7 responses, the state space models have maintained their superiority in terms of $Adj - R^2$ and the SE of Regression. Thus the state space models provide better explanatory power in terms of higher R^2 and lower error variances of regression.

Next we examine the significance and stability of the new updated coefficients of X_5 , X_6 and X_7 , since these are the main coefficients to be used as inputs for the construction of our mean-variance frontier for the $MI - Garch(1, 1)$. The reader is referred to Appendix

4A, under the 'List of Appendices' section of this study for the eviws output concerning these coefficients.

Scrutinising the $MI - GARCH(1, 1)$ coefficients for the 9 responses, it is clear that they are all non-zero. As seen before with our MI-time series, regression and state space models, the subset of explanatory variables (X_5, X_6, X_7) is not always significant due to the problem of fitting on a common platform. Thus again it is clear in the $GARCH(1, 1)$ scenario, that all the models do not always exhibit highly significant coefficients for this subset all the time. Nonetheless, the coefficients attained for this subset is non-zero, which is important for our mean-variance construction.

6.13.5 Conclusions

Having completed implementing the $MI - GARCH(1, 1)$ methodologies on the respective responses, the following is concluded: that significant $MI - GARCH(1, 1)$ models have been found for all 9 stocks of our portfolio. Furthermore the $MI - GARCH(1, 1)$ models acquired appear to provide better explanatory power in terms of high R^2 values and low error variances of the models than the state space models for the stocks R_4 and R_2 .. Furthermore we conclude that State Space models are more superior with regards to a modelling sense. But it was the $MI - GARCH(1, 1)$ models that demonstrated superior beta estimates and values compared to state space models. Hence it still remains to be seen which is superior with regards to the mean-variance frontier. The betas and regression SE's computed for the 9 stocks will be used as inputs for the computation of the efficient frontier for the single index state space model. The variance JSE and the mean log returns

6.14 List of Appendices

6.14.1 Appendix 4A: MI-Garch(1,1) Coefficients

	Coefficient	Std. Error	z-Statistic	Prob.
C	-0.004228	0.003130	-1.350595	0.1768
X5	0.131779	0.030656	4.298589	0.0000
X6	1.255178	0.061681	20.34950	0.0000
X7	0.033070	0.024008	1.377436	0.1684
AR(2)	-0.164404	0.066193	-2.483697	0.0130

R1 - coefficients

	Coefficient	Std. Error	z-Statistic	Prob.
C	0.010449	0.004233	2.468458	0.0136
X5	-0.309728	0.043718	-7.084644	0.0000
X6	1.085865	0.090800	11.95892	0.0000
X7	0.029880	0.054050	0.552819	0.5804
AR(1)	-0.329169	0.089601	-3.673732	0.0002

R3 - coefficients

	Coefficient	Std. Error	z-Statistic	Prob.
@SQRT(GARCH)	-0.872255	1.398583	-0.623671	0.5328
C	0.078690	0.108679	0.724061	0.4690
X5	-0.084471	0.054411	-1.552462	0.1206
X6	0.414302	0.148480	2.790287	0.0053
X7	-0.011747	0.048748	-0.240981	0.8096

R5 - Coefficients

	Coefficient	Std. Error	z-Statistic	Prob.
C	0.005216	0.006248	0.834750	0.4039
X5	-0.073194	0.052285	-1.399890	0.1615
X6	1.030093	0.125013	8.239904	0.0000
X7	0.090766	0.049227	1.843829	0.0652

R7 - Coefficients

	Coefficient	Std. Error	z-Statistic	Prob.
C	1.66E-05	0.004795	0.003454	0.9972
X5	-0.065516	0.050868	-1.287959	0.1978
X6	0.766652	0.140437	5.459061	0.0000
X7	0.128435	0.049691	2.584689	0.0097
AR(2)	-0.202761	0.060981	-3.324972	0.0009

R9 - Coefficients

	Coefficient	Std. Error	z-Statistic	Prob.
@SQRT(GARCH)	-17.31288	46.37692	-0.373308	0.7089
C	2.047923	5.483687	0.373457	0.7088
X5	-0.199493	0.081848	-2.437345	0.0148
X6	1.367439	0.126712	10.79169	0.0000
X7	-0.043810	0.088281	-0.496250	0.6197

R2 - Coefficients

	Coefficient	Std. Error	z-Statistic	Prob.
@SQRT(GARCH)	6.033323	10.71682	0.562977	0.5735
C	-0.344508	0.625069	-0.551152	0.5815
X5	-0.078924	0.037548	-2.101936	0.0356
X6	0.905099	0.090571	9.993228	0.0000
X7	-0.032774	0.035561	-0.921634	0.3567
AR(1)	-0.190852	0.070564	-2.704659	0.0068

R4 - Coefficients

	Coefficient	Std. Error	z-Statistic	Prob.
LOG(GARCH)	0.018284	0.017591	1.039391	0.2986
C	0.087162	0.090105	0.967334	0.3334
X5	-0.013398	0.045996	-0.291289	0.7708
X6	0.902363	0.133762	6.746054	0.0000
X7	0.212402	0.054952	3.865235	0.0001

R6 - Coefficients

	Coefficient	Std. Error	z-Statistic	Prob.
LOG(GARCH)	0.010342	0.016729	0.618224	0.5364
C	0.060582	0.092398	0.655660	0.5120
X5	-0.167134	0.042281	-3.952905	0.0001
X6	0.939854	0.090744	10.35719	0.0000
X7	-0.031467	0.042968	-0.732328	0.4640
AR(6)	-0.228132	0.071448	-3.192972	0.0014

R8 - Coefficients

6.15 A Summary of the ARCH and GARCH Models

The general $ARMA(p, q)$ model can be written in the form

$$r_t = \phi_0 + \sum_{i=1}^p \phi_i r_{t-i} - \sum_{i=1}^q \theta_i e_{t-i} + e_t \quad (1)$$

$$= \mu_t + e_t \quad (2)$$

where e_t is a white noise series $E(e_t/F_{t-1}) = 0$, $var(e_t/F_{t-1}) = \sigma^2$ and F_{t-1} is the conditional state of nature at time $t - 1$. Financial Times Series Models, especially Stock Exchange Returns (log) very seldom have the above property. More often, the model is of the form

$$r_t = \mu + e_t \quad (3)$$

where μ is a constant and the returns, r_t , that is the errors e_t , (a) are serially uncorrelated but depended and (b) the dependence of e_t can be described by a simple quadratic function of the lagged (past) values of e_t , that is e_{t-1}, e_{t-2}, \dots . Specifically, an $ARCH(m)$ assumes that

$$r_t - \mu = e_t = \sigma_t \varepsilon_t$$

$$\sigma_t^2 = \alpha_0 + \alpha_1 e_{t-1}^2 + \dots + \alpha_m e_{t-m}^2$$

where $\{\varepsilon_t\}$ is a series of identically and independently (i.i.d.) random variables with mean zero and variance 1 and $\alpha_i \geq 0$ for $i > 0$. This is clearly an auto-regressive conditional

heteroskedastic model since

$$\begin{aligned}
 \text{var}(r_t|F_{t-1}) &= E(r_t - \mu)^2|F_{t-1}) \\
 &= E(\sigma_t \epsilon|F_{t-1})^2 \\
 &= \sigma_t^2. \\
 &= \alpha_0 + \alpha_1 e_{t-1}^2 + \cdots + \alpha_m e_{t-m}^2.
 \end{aligned}$$

In practice, ϵ_t is often assumed to have a normal distribution. From the structure of the model, it is seen that large past squared shocks $\{e_t^2\}_{t=1}^m$ imply a large conditional variance σ_t^2 for the return r_t and e_t . Consequently e_t tends to assume a large value (in modulus). This means, that under the *ARCH* framework, large shocks tend to be followed by another large shock. This feature is similar to the volatility clustering observed in most asset returns.

We can estimate the parameters of an *ARCH*(m) model by using the conditional log likelihood function. In general we assume that ϵ_t is either a standard normal variate or that it has a standardised Students t distribution. If ϵ_t is normally distributed then the conditional log likelihood is:

$$\log \left\{ \prod_{t=m+1}^T \frac{1}{\sqrt{2\pi\sigma_t^2}} \exp(-e_t^2/\sigma_t^2) \right\} = \sum_{t=p+1}^T \left\{ -\frac{1}{2} \log(2\pi) - \frac{1}{2} \log(\sigma_t^2) - \frac{1}{2} \frac{e_t^2}{\sigma_t^2} \right\}, \quad (4)$$

where $\sigma_t^2 = \alpha_0 + \alpha_1 e_{t-1}^2 + \alpha_2 e_{t-2}^2 + \dots + \alpha_p e_{t-p}^2$. Equation 4.5 should be evaluated iteratively for each observation in order to maximise the conditional log likelihood function. It can be seen that $e_t \sim N(0, \sigma_t^2)$. Bollerslev (1986) developed the Generalized *ARCH* models since it was found that often *ARCH* models required relatively long lag structures in the conditional variance equation. The *GARCH* model is an extension of the univariate

ARCH similar to the way in which *ARMA* models are extensions of *AR* and *MA* models.

The model can be defined as follows:

Let r_t be the return (log) of a share at time t and μ_t is the mean of a share at time t .

It is then said that R_t follows a *GARCH*(p, q) process if and only if:

$$\begin{aligned} r_t - \mu_t &= e_t = \sigma_t \varepsilon_t \quad \text{with} \\ \sigma_t^2 &= \alpha_0 + \sum_{i=1}^p \alpha_i e_{t-i}^2 + \sum_{j=1}^q \beta_j \sigma_{t-j}^2 \end{aligned} \quad (5)$$

where ε_t is a white noise random variable such that $E(\varepsilon_t) = 0$ and $Var(\varepsilon_t) = 1$. Notice that in this formulation p represents the number of *ARCH* terms where as q represents the number of *GARCH* terms.

6.16 Additional ARCH Models

6.16.1 The EGARCH Model

Black (1976) and Christie (1982) both found that there exists a negative correlation between current returns and future volatility of share returns. i.e. volatility tends to rise in response to bad news and fall with good news (where good news indicates the situation in which returns are larger than the consensus view). The standard *GARCH* model cannot capture this observation since it models the conditional volatility as the sum of squared lagged returns and lagged conditional variances. The size of the returns at each time period is thus more important than the sign in a *GARCH* context. Nelson (1991) developed the exponential *GARCH* (*EGARCH*) in order to capture the above observation. The

conditional variance is modelled as:

$$\log(h_t) = \alpha_t + \sum_{j=1}^{\infty} \beta_j g(\varepsilon_{t-j}) \quad (1)$$

where $\{\alpha_t\}_{t=-\infty, \infty}$ and $\{\beta_j\}_{j=1, \infty}$ are real, non-stochastic scalars, while $g(\cdot)$ is a function of both the size and sign of the return during time period t . Nelson (1991) proposed using:

$$g(\varepsilon_t) = \theta \varepsilon_t + \gamma [|\varepsilon_t| - E(|\varepsilon_t|)] \quad (2)$$

and thus:

$$g(\varepsilon_t) = \begin{cases} (\theta + \gamma) \varepsilon_t - E(|\varepsilon_t|) & \text{for } \varepsilon_t > 0 \\ (\theta - \gamma) \varepsilon_t - E(|\varepsilon_t|) & \text{for } \varepsilon_t \leq 0 \end{cases} \quad (3)$$

It can be seen that $g(\cdot)$ models the conditional variance asymmetrically and that it does allow for a negative relationship to exist between returns and future conditional volatility. (i.e. if $\gamma = 0$ and $\theta < 0$ then the change in the conditional variance is positive (negative) when $\varepsilon_t < 0$. ($\varepsilon_t > 0$). The GED (Generalized Error Distribution (Harvey 1981), Box and Tiao (1973)) is often used to model the innovation distribution. It is defined as follows:

$$f(x) = \frac{v \exp\left[\frac{-1}{2} \left|\frac{x}{\lambda}\right|^v\right]}{\lambda 2^{(1+\frac{1}{v})} \Gamma(\frac{1}{v})} \quad \text{where } -\infty < x < \infty \quad (4)$$

where $\lambda = \sqrt{2^{\frac{(-2)}{v}} \frac{\Gamma(\frac{1}{v})}{\Gamma(\frac{3}{v})}}$, $\Gamma(\cdot)$ is the gamma function and $v > 0$ is a tail parameter to be estimated. The GED contains the normal distribution as a special case (i.e. when $v = 2$), however heavy tailed distributions can be modelled by setting $v < 2$. Similarly thinner tailed distributions (than the normal) is modelled by setting $v > 2$

6.16.2 The Threshold GARCH (TARCH) Model

TARCH or Threshold *ARCH* and Threshold *GARCH* were introduced independently by Zakoian (1994) and Glosten, Jaganathan, and Runkle (1993). The generalized specification for the conditional variance is given by

$$h_t = \sigma_t^2 = \omega + \sum_{j=1}^q \beta_j \sigma_{t-j}^2 + \sum_{i=1}^p \alpha_i \varepsilon_{t-i}^2 + \sum_{k=1}^r \gamma_k \varepsilon_{t-k}^2 \Gamma_{t-k} \quad (1)$$

where $\Gamma_t = 1$ if $\varepsilon_t < 0$ and 0 otherwise. In this model, good news $\varepsilon_{t-i} > 0$, and bad news $\varepsilon_{t-i} < 0$, have differential effects on the conditional variance; good news has an impact of α_i , while bad news has an impact of $\alpha_i + \gamma_i$. If $\gamma_i > 0$, bad news increases volatility, and we can say that there is a leverage effect for the i -th order. If $\gamma_i \neq 0$, the news impact is asymmetric.

Note that *GARCH* is a special case of the *TARCH* model where the threshold term is set to zero. To estimate a TARCH model, we specify our *GRACH* model with *ARCH* and *GARCH* order and then change the Threshold order to the desired value.

6.16.3 The Power ARCH (PARCH) Model

Taylor (1986) and Schwert (1989) introduced the standard deviation GARCH model, where the standard deviation is modelled rather than the variance. This model along with several other models is generalised in Ding *et al.* (1993) with the Power *ARCH* specification. In the Power *ARCH* model, the power parameter δ of the standard deviation can be estimated rather than imposed, and the optional γ parameters are added to capture asymmetry of up

to order τ :

$$\sigma_t^\delta = \omega + \sum_{j=1}^q \beta_j \sigma_{t-j}^\delta + \sum_{i=1}^p \alpha_i (|\varepsilon_{t-i}| - \gamma_i \varepsilon_{t-i}^2)^\delta \quad (1)$$

where $\delta > 0$, $|\gamma_i| \leq 1$ for $i = 1, \dots, \tau$, $\gamma_i = 0$ for all $i > \tau$, and $\tau \leq p$. The symmetric model sets $\gamma_i = 0$ for all i . Note that if $\delta = 2$ and $\gamma_i = 0$ for all i , the *PARCH* model is simply a standard *GARCH* specification. Akin the previous models, the asymmetric effects are present if $\gamma \neq 0$.

6.16.4 The Component GARCH (C-GARCH) Model

The conditional variance in the *GARCH*(1, 1) model:

$$\sigma_t^2 = \bar{\omega} + \alpha (\varepsilon_{t-1}^2 - \bar{\omega}) + \beta (\sigma_{t-1}^2 - \bar{\omega}) \quad (1)$$

shows mean reversion to $\bar{\omega}$, which is a constant for all time. By contrast, the component *GARCH* model allows mean reversion to a varying level m_t , modelled as:

$$\begin{aligned} \sigma_t^2 - m_t &= \bar{\omega} + \alpha (\varepsilon_{t-1}^2 - \bar{\omega}) + \beta (\sigma_{t-1}^2 - \bar{\omega}) \\ m_t &= \omega + \rho(m_{t-1} - \omega) + \phi(\varepsilon_{t-1}^2 - \sigma_{t-1}^2) \end{aligned} \quad (2)$$

Here σ_t^2 is still the volatility, while m_t takes the place of ω and is the time varying long run volatility. The first equation describes the transitory component, $\sigma_t^2 - m_t$, which converges to zero with powers of $(\alpha + \beta)$. The second equation describes the long run component m_t , which converges to ω with powers of ρ . ρ is typically between 0.99 and 1 so that m_t approaches ω very slowly. We can combine the transitory and permanent equations and

write:

$$\begin{aligned}\sigma_t^2 = & (1 - \alpha - \beta)(1 - \rho)\omega + (\alpha + \phi)\varepsilon_{t-1}^2 - (\alpha\rho + (\alpha + \beta)\phi)\varepsilon_{t-2}^2 \\ & + (\beta - \phi)\sigma_{t-1}^2 - (\beta\rho - (\alpha + \beta)\phi)\sigma_{t-2}^2\end{aligned}\quad (3)$$

which shows that the component model is a (nonlinear) restricted *GARCH*(2, 2) model.

6.17 Empirical Investigation: Single Index

6.17.1 The data

For this empirical analysis, the data used was the SIDS batch.

6.17.2 Study Objectives

Thus far single index models have been established for the 9 stocks in our portfolio in the regression, time series, state space and *GARCH*(1, 1) contexts. For all 4 modelling scenarios, the relevant beta coefficients of the JSE and SE of regressions have been collected for our portfolio in all 4 modelling environs which are going to be used to finally construct the respective mean-variance frontiers for each modelling scenario. In addition we have conducted model comparisons between the 4 modelling paradigms up to now.

One of the outlined objectives in this thesis is to show the supremacy of *GARCH*(1, 1) over its various extensions, namely *E-GARCH*, *PARCH*, *TARCH* and *C-GARCH*. Thus the time series and regression models found initially in the earlier studies are now going to be re-constructed under the 4 *GARCH* extensions. The objective in this study is to

extract and compare the relevant Beta coefficients and SE of regressions which will assist in the final construction of efficient frontiers of the 4 *GARCH* extensions.

As a consequence, the purpose of this study is to construct the 4 additional *GARCH* extensions in Eviews 5.1 and compare the extensions in terms of $Adj - R^2$, their SE of Regression and finally the stability of their beta coefficients with that of our simple *GARCH*(1,1).

6.17.3 Methodology

There are 3 steps when building any type of *ARCH* Model:

1. First one needs to establish an econometric model for the return series in order to eliminate any linear dependence in the data , e.g. a time series or regression model - *ARMA*(1,1) for example. Then use the residual series of the model to test for *ARCH* effects.
2. Then one needs to specify the *ARCH* order and then execute the estimation process
3. The one needs to specify a threshold order (if it is a *TARCH* model). Otherwise one needs to specify the asymmetric order if it happens to be an *EGARCH* model. In the case of a *PARCH* model, the analyst needs to fix a power parameter with Asymmetric order 1. Finally for the *C - GARCH* model, the analyst has the option to include a threshold term, as long as it adds to the mean reverting effect of the model.
4. Finally an empirical and significance check of the various extension-*ARCH* model fit needs to be reviewed carefully and then to refine it if necessary.

From chapter 4, we showed that the response variables R_1, R_3, R_4, R_8 and R_9 had AR time series components, while the rest were in the form of simple regression models accompanied by only one explanatory, JSE. These 9 models found under the time series context will now be taken in their respective forms and refitted with their explanatory and AR components under the various $GARCH$ extensions.

Once fitting the respective models for all nine responses, the model selection procedure is based on the R^2 , Adjusted R^2 and the SE of Regression. Once the appropriate $GARCH$ extensions are fit for the respective responses, the most efficient model is opted for based on the aforementioned selection criteria. Subsequent to the model selection procedure we will conclude with a comparison between $GARCH(1, 1)$ and its various extensions.

6.17.4 Primary Findings

Implementing the above methodology for $E - GARCH$, $PARCH$, $C - GARCH$ and $TARCH$, the following statistics concerning their model selection criteria are highlighted in table 6.7 below:

Responses	E-GARCH		C-GARCH		PARCH		TARCH		GARCH(1,1)	
	Adj- R^2	S.E. of reg	Adj- R^2	S.E. of reg	Adj- R^2	S.E. of reg	Adj- R^2	S.E. of reg	Adj- R^2	S.E. of reg
Model for R1 (Model 1)	0.6782	0.0566	0.6746	0.0570	0.6778	0.0567	0.6793	0.0566	0.6824	0.0563
Model for R2 (Model 2)	0.1637	0.1254	0.1801	0.1241	0.1649	0.1253	0.1786	0.1242	0.1807	0.1241
Model for R3 (Model 3)	0.1893	0.0920	0.1904	0.0919	0.1879	0.0921	0.1922	0.0918	0.1959	0.0916
Model for R4 (Model 4)	0.3897	0.0621	0.3950	0.0619	0.3926	0.0620	0.3943	0.0619	0.4534	0.0599
Model for R5 (Model 5)	0.1037	0.0793	0.0222	0.0829	0.0459	0.0833	0.0111	0.0833	0.0714	0.0820
Model for R6 (Model 6)	0.3081	0.0921	0.3154	0.0916	0.3096	0.0920	0.3094	0.0920	0.3119	0.0919
Model for R7 (Model 7)	0.3476	0.0797	0.3483	0.0796	0.3469	0.0797	0.3502	0.0795	0.3559	0.0792
Model for R8 (Model 8)	0.2785	0.0661	0.2743	0.0662	0.2670	0.0666	0.2788	0.0660	0.2825	0.0659
Model for R9 (Model 9)	0.2127	0.0815	0.2084	0.0817	0.2071	0.0818	0.2154	0.0813	0.2193	0.0811

Table 6.7: Selection Criteria Stats for GARCH(1,1) vs GARCH Extensions

As expected and illustrated by table 6.7, the $GARCH(1, 1)$ has out performed it's 4 ex-

tensions for all 9 response variables in terms of achieving a higher Adjusted R^2 and lower error variances. Thus the $GARCH(1, 1)$ model explains the most variation for all 9 responses of the other 4 $GARCH$ extensions and hence provides a better fit of the data. A summary table of β estimates and their respective p-values are presented below in table

6.8:

Responses	E-GARCH		C-GARCH		PARCH		TARCH		GARCH(1,1)	
	JSE β	p-value	JSE β	p-value	JSE β	p-value	JSE β	p-value	JSE β	p-value
Model for R1 (Model 1)	1.5058	0.0000	1.4845	0.0000	1.5444	0.0000	1.4602	0.0000	1.4735	0.0000
Model for R2 (Model 2)	0.8722	0.0000	0.9263	0.0000	0.9374	0.0000	1.0032	0.0000	1.1383	0.0000
Model for R3 (Model 3)	0.6654	0.0000	0.6234	0.0000	0.6099	0.0000	0.6055	0.0000	0.6320	0.0000
Model for R4 (Model 4)	0.8174	0.0000	0.7696	0.0000	0.7914	0.0000	0.8156	0.0000	0.8285	0.0000
Model for R5 (Model 5)	1.1981	0.0007	0.2548	0.0045	0.2588	0.0007	0.2746	0.0005	0.2647	0.0006
Model for R6 (Model 6)	0.9996	0.0000	0.9540	0.0000	1.1163	0.0000	1.0312	0.0000	1.0014	0.0000
Model for R7 (Model 7)	0.9602	0.0000	0.9557	0.0000	0.9571	0.0000	0.9799	0.0000	0.9803	0.0000
Model for R8 (Model 8)	0.7177	0.0000	0.7193	0.0000	0.7225	0.0000	0.7366	0.0000	0.7424	0.0000
Model for R9 (Model 9)	0.7497	0.0000	0.7050	0.0000	0.7585	0.0000	0.7481	0.0000	0.7477	0.0000

Table 6.8: p-values and β estimates for single index model - GARCH(1,1) vs

Extensions

Observing table 6.8 above, it is clear that all the β estimates are significant at the 5% level of significance. Overall, the $GARCH(1, 1)$ model exhibits more stable β estimates, this can be attributed to the lower standard errors achieved by the model.

6.17.5 Conclusions

It is concluded that the $GARCH(1, 1)$ model still exhibits the greater explanatory power and supreme model fit for our financial data. In addition the $GARCH(1, 1)$ model provides significantly more stable β estimates and lower error variances in a modelling context.

Having constructed the 4 extensions of $GARCH(1, 1)$ and extracted their respective SE-of regressions and β' s, we are now left with the vital task of constructing their respective efficient frontiers, in order to determine which modelling technique is in fact the more

superior in terms of a mean-variance (risk) framework. This follows in the subsequent chapter.

Chapter 7

Modern Portfolio Theory

7.1 Classical Formulations

We now develop a portfolio model. In our portfolio we have a vector of p stocks. In general p is large and can be very large. Portfolios held by Pension and Mutual Funds are large indeed. On the JSE alone there are well over a 1000 stocks. Let the vector of stocks be

$$R = \begin{pmatrix} R_1 \\ \vdots \\ R_p \end{pmatrix} \quad (1)$$

with the expected returns

$$E(R) = \mu. \quad (2)$$

Assume that the covariance between returns of different shares are non-zero thus the covariance matrix of the stock returns is

$$\Sigma = E(R - \mu)(R - \mu)'. \quad (3)$$

$$\Sigma = \begin{pmatrix} \sigma_{11} & \dots & \sigma_{1p} \\ \vdots & & \vdots \\ \sigma_{p1} & \dots & \sigma_{pp} \end{pmatrix} \quad (4)$$

Where $\sigma_{ii} = \sigma_i^2$ is the variance of the i 'th stock and σ_{ij} the covariance between the i 'th and the j 'th stock. Further assume that

$$R \sim N(\mu, \Sigma) \quad (5)$$

that is multivariate normal. A portfolio of stocks can be thought of as a cash investment, say w_i in each stock. We can assume that this w_i is a proportion of the wealth available to us, so that

$$\sum_{i=1}^p w_i = 1. \quad (6)$$

Let

$$\mathbf{W} = \begin{pmatrix} w_1 \\ \vdots \\ w_p \end{pmatrix} \quad (7)$$

then our portfolio is

$$\begin{aligned} P &= \mathbf{W}'\mathbf{R} \\ &= \sum_{i=1}^p w_i R_i \end{aligned} \quad (8)$$

subject to

$$\sum_{i=1}^p w_i = 1.$$

The expected return of the portfolio is

$$\begin{aligned} E(P) &= \mathbf{W}'E(\mathbf{R}) \\ &= \mathbf{W}'\boldsymbol{\mu} \end{aligned} \quad (9)$$

$$= \sum_{i=1}^p w_i \mu_i \quad (10)$$

$$= \mu_p \quad (11)$$

and the variance is

$$\begin{aligned} \text{var}(P) &= \mathbf{W}'\boldsymbol{\Sigma}\mathbf{W} \\ &= \sum_{i=1}^p \sum_{j=1}^p w_i w_j \sigma_{ij} = \sigma_p^2 \end{aligned}$$

If in addition $R \sim N(\mu, \Sigma)$ then $P \sim N(\mu_p, \sigma_p^2)$. By changing the weights w_i , this will also impose a change in the value of the portfolio and its expected values. Clearly we want to choose the weights in order to provide us as large a possible expected return $E(P) = \mu_p$. Thus we want maximize the expected return $E(P) = \mu_p$. But this will not help us much if the variance of the portfolio σ_p^2 is also large. In Finance we often interpret the variance of the portfolio as its 'risks'. The larger the variance the larger the risk and vice versa. Thus we want to choose the weights w_i such that the expected return $E(P) = \mu_p$ is a maximum but also at the same time that the risk or variance σ_p^2 is a minimum. Our portfolio problem is then

$$\max_{w_i} E(P) = \mathbf{W}'\boldsymbol{\mu} \quad (12)$$

$$= \sum_{i=1}^p w_i \mu_i$$

$$\min_{w_i} \sigma_p^2 = \mathbf{W}'\boldsymbol{\Sigma}\mathbf{W} \quad (13)$$

$$= \sum_{i,j=1}^p w_i w_j \sigma_{ij} \quad (14)$$

subject to

$$\sum_{i=1}^p w_i = 1.$$

This formulation is due to Harry Markowitz. Markowitz, added to this the concept of an Efficient Frontier.

7.1.1 Efficient Frontier.

Definition1 A portfolio is called efficient if:

- For a given amount of risk, the expected return is maximized, or:
- For a given amount of return, the expected risk is minimized, and:
- The portfolio must be legitimate (no negative sales).

We then generate the Efficient Frontier and the investor takes a position somewhere on this frontier. The Markowitz portfolio problem is then formally:

$$\max_{w_i} E(P) = \mathbf{W}'\boldsymbol{\mu} \quad (15)$$

$$= \sum_{i=1}^p w_i \mu_i$$

$$\min_{w_i} \sigma_p^2 = \mathbf{W}'\boldsymbol{\Sigma}\mathbf{W} \quad (16)$$

$$= \sum_{i,j=1}^p w_i w_j \sigma_{ij} \text{ subject to}$$

$$\sum_{i=1}^p w_i = 1, \quad (17)$$

$$0 \leq w_i \leq 1, \quad i = 1, \dots, p. \quad (18)$$

This is a non-linear (quadratic) programming problem called a *QP – problem* and can now be solved using well-known numerical techniques. Markowitz was awarded the Nobel prize, in Economics and Finance, for the above formulation. To find the Efficient Frontier is a formidable task. It is problem of constrained quadratic optimization (*QP*). To solve the problem one needs either to fix the return μ_p and then minimize the variance σ_p^2 , or fix the variance and then maximize the return. They will both lead to the same answer. We

prefer to fix the return. Thus we need to solve the following constrained QP problem.

$$\begin{aligned}
 \text{Min } \sigma_p^2 &= \mathbf{W}'\Sigma\mathbf{W} \\
 &= \sum_{i,j=1}^p w_i w_j \sigma_{ij} \text{ subject to} \\
 \mu_p &= \mathbf{W}'\boldsymbol{\mu}_p \\
 &= \sum_{i=1}^p w_i \mu_i \tag{19}
 \end{aligned}$$

$$= E_k \quad \text{subject to} \tag{20}$$

$$\sum_{i=1}^p w_i = 1, \quad 0 \leq w_i \leq 1,$$

where E_k is a fixed expected return on the μ_p (Y) axis. By varying E_k a sufficient number of times will generate the efficient frontier. For the equality constraints we can use Lagrange multipliers. Thus the constrained QP problem becomes

$$\begin{aligned}
 \text{Min } Z &= \mathbf{W}'\Sigma\mathbf{W} - \phi \left(\sum_{i=1}^p w_i \mu_i - E_k \right) \\
 &\quad - \lambda \left(\sum_{i=1}^p w_i - 1 \right) \text{ subject to} \\
 0 &\leq w_i \leq 1, \quad i = 1, \dots, p
 \end{aligned} \tag{21}$$

Varying E_k will yield the Efficient Frontier. This is a standard QP problem with bounds $0 \leq w_i \leq 1$, $i = 1, \dots, p$ and algorithms are available on most computers including PC's. In fact more general constraints, equality as well as inequality constraints, can be handled by such a QP problem. We will discuss this further when solving the modern portfolio problem. An example of different bounds would be

$$L_i \leq w_i \leq U_i, \quad i = 1, \dots, N, \tag{22}$$

where L_i is a lower bound imposed on a stock (buy at least a proportion L_i of a stock), and U_i is an upper bound imposed on a stock (do not buy more than a proportion U_i of the stock). Such imposed bounds are very common and often used. Note we still have the constraint that $L_i \geq 0$. Although QP algorithms are now readily available it did not use to be the case. Some huge computer companies had a monopoly over the algorithms and made it only available to their own clients or else with a huge fee.

Sharpe introduced the concept of corner portfolios which reduced the QP problem to a LP problem. We will now develop modern portfolio theory which will include most of Sharpe's development.

7.2 Modern Portfolio Theory: The Corner Portfolios of Sharpe.

Consider the Efficient Frontier for the 9 stocks. We draw a straight line through the frontier.

Call this line

$$Y = A + BX \text{ or } \mu_p = A + B\sigma_p^2 \quad (1)$$

$$A = Y - BX \text{ or } A = \mu_p - B\sigma_p^2$$

$$\frac{A}{B} = \frac{1}{B}Y - X \text{ or } \frac{A}{B} = \frac{1}{B}\mu_p - \sigma_p^2$$

$$Z = \phi Y - X \text{ or } Z = \phi\mu_p - \sigma_p^2.$$

Let the slope

$$B = \frac{1}{\phi} \quad (2)$$

then the line parallel to the $\sigma_p^2 = X$ axis is where the slope $B = 0$, that is $\phi = \infty$, and the line parallel to the $\mu_p = Y$ axis is the line where the slope $B = \infty$ or $\phi = 0$. If we now take the line

$$\frac{A}{B} = \frac{1}{B}\mu_p - \sigma_p^2 \text{ or } Z = \phi\mu_p - \sigma_p^2. \quad (3)$$

and maximize A keeping B fixed, that is maximize Z keeping ϕ fixed will give us a point on the Efficient Frontier. Now by varying the slope B that is ϕ will generate the Efficient Frontier. This is the critical line algorithm subject to the portfolio constraint $\sum_{i=1}^p w_i = 1$. Thus our problem reduces to

$$\begin{aligned} \text{Max } Z &= \phi\mu_p - \sigma_p^2 \\ &= \phi \mathbf{W}'\boldsymbol{\mu} - \mathbf{W}'\boldsymbol{\Sigma}\mathbf{W} \text{ subject to} \\ \sum_{i=1}^p w_i &= 1. \end{aligned} \quad (4)$$

We vary ϕ from $(0, \infty)$ to generate the Efficient Frontier. Note we have made no other demands on the $w_i, i = 1, \dots, p$ except that

$$\sum_{i=1}^p w_i = 1.$$

The fact that w_i is not constrained to take on only positive values means that some of the w_i can take on negative values. This implies leverage (selling a stock while not having it - called short selling) while w_i positive means buying the stock. In all stock exchanges over the world this is now a common trading procedure. When Markowitz and Sharpe and others developed portfolio theory, this trading procedure was considered to be not legitimate. Thus in modern portfolio theory buying and selling are normal strategies. It

could lead to enormous profits, or hornworms losses. The riskiest job in the financial area is that of portfolio manager.

We now give an analytic solution to the portfolio problem formulated in (4). The portfolio constraints $\sum_{i=1}^p w_i = 1$ can be included in the objective function (4) using Lagrange multipliers as follows

$$\begin{aligned} \text{Max } Z' &= \phi \mathbf{W}' \boldsymbol{\mu} - \mathbf{W}' \boldsymbol{\Sigma} \mathbf{W} + \lambda \left(1 - \sum_{i=1}^p w_i\right) \\ &= \phi \sum_{i=1}^p w_i \mu_i - \sum_{i=1}^p \sum_{j=1}^p w_i w_j \sigma_{ij} + \lambda \left(1 - \sum_{i=1}^p w_i\right) \end{aligned} \quad (5)$$

Taking partial derivatives of Z' with respect to $w_i, i = 1, \dots, p$ and λ and setting them to zero gives the simultaneous equations

$$\begin{aligned} \frac{\partial Z'}{\partial w_1} &= \phi \mu_1 - 2w_1 \sigma_{11} - 2w_2 \sigma_{12} - \dots - 2w_p \sigma_{1p} - \lambda = 0 \\ \frac{\partial Z'}{\partial w_2} &= \phi \mu_2 - 2w_2 \sigma_{22} - 2w_1 \sigma_{21} - \dots - 2w_p \sigma_{2p} - \lambda = 0 \\ &\vdots \\ \frac{\partial Z'}{\partial w_p} &= \phi \mu_p - 2w_p \sigma_{pp} - 2w_1 \sigma_{p1} - \dots - 2w_{p-1} \sigma_{p,p-1} - \lambda = 0 \\ \frac{\partial Z'}{\partial \lambda} &= 1 - w_1 - w_2 - \dots - w_p = 0 \end{aligned} \quad (6)$$

or in matrix notation

$$\begin{pmatrix} 2\sigma_{11} & 2\sigma_{12} & \dots & 2\sigma_{1p} & 1 \\ 2\sigma_{21} & 2\sigma_{22} & \dots & 2\sigma_{2p} & 1 \\ \vdots & \vdots & \vdots & \vdots & \vdots \\ 2\sigma_{p1} & 2\sigma_{p2} & \dots & 2\sigma_{pp} & 1 \\ 1 & 1 & \dots & 1 & 0 \end{pmatrix} \cdot \begin{pmatrix} w_1 \\ w_2 \\ \vdots \\ w_p \\ \lambda \end{pmatrix} = \begin{pmatrix} \phi \mu_1 \\ \phi \mu_2 \\ \vdots \\ \phi \mu_p \\ 1 \end{pmatrix} \quad (7)$$

or

$$\mathbf{AX} = \mathbf{B}$$

with solution

$$\mathbf{X} = \mathbf{A}^{-1}\mathbf{B}. \quad (8)$$

The solutions are of the following form

$$w_i = c_i + d_i(\phi), \quad i = 1, \dots, p, \quad (9)$$

$$\lambda = c_\lambda + d_\lambda(\phi).$$

By varying ϕ from zero to infinity, the set of weights which is optimum for each value of ϕ is generated. For any specific value of ϕ the portfolio mean and variance is then

$$\begin{aligned} E(P) &= \mu_p = \mathbf{W}'\boldsymbol{\mu} \\ &= \sum_{i=1}^p w_i \mu_i \\ &= \sum_{i=1}^p (c_i + d_i(\phi)) \mu_i \end{aligned} \quad (10)$$

$$= \sum_{i=1}^p c_i \mu_i + \phi \sum_{i=1}^p d_i \mu_i \quad (11)$$

with variance

$$var(P) = \sigma_p^2 = \mathbf{W}'\boldsymbol{\Sigma}\mathbf{W} \quad (12)$$

$$\begin{aligned} &= \sum_{i=1}^p \sum_{j=1}^p w_i w_j \sigma_{ij} \\ &= \sum_{i=1}^p \sum_{j=1}^p \{c_i + d_i(\phi)\} \{c_j + d_j(\phi)\} \\ &= \sum_{i=1}^p \sum_{j=1}^p c_i c_j + \phi \sum_{i=1}^p \sum_{j=1}^p (c_i d_j + c_j d_i) \\ &\quad + \phi^2 \sum_{i=1}^p \sum_{j=1}^p d_i d_j. \end{aligned} \quad (13)$$

It should be noted that some weights can become negative. This is called leverage (short selling). It may be against the policy decision of a portfolio manager to use leverage (short selling). The above simple calculus procedure can still be used to generate legitimate portfolios. As soon as the first stock's weight reaches zero (before it becomes negative) this stock is eliminated from the portfolio- that is the corresponding row and column of the bordered covariance matrix in equation (7) is eliminated and the procedure is started again.

For such portfolios it is probably easier to use the QP algorithm with bounds. Alternatively Sharpe's LP algorithm using corner portfolios could be used. Before discussing Sharpe's simplified solution we would like to discuss the addition of further constraints to the portfolio problem given in (4). As long as the constraints are equality constraints they can easily be added to the calculus problem using Lagrange multipliers. As an example suppose that several equality constraints of following type are added. Adding constraints:

$$\begin{aligned}
 d_{11}w_1 + d_{12}w_2 + \dots + d_{1p}w_p &= R_1 \\
 d_{21}w_1 + d_{22}w_2 + \dots + d_{2p}w_p &= R_2 \\
 &\vdots = \vdots \\
 d_{m1}w_1 + d_{m2}w_2 + \dots + d_{mp}w_p &= R_m.
 \end{aligned} \tag{14}$$

Now maximize

$$\begin{aligned}
 Z' = & \phi\mu_p - \sigma_p^2 \\
 & + \lambda_1(R_1 - d_{11}w_1 - d_{12}w_2 - \dots - d_{1p}w_p) \\
 & + \lambda_2(R_2 - d_{21}w_1 - d_{22}w_2 - \dots - d_{2p}w_p) \\
 & + \dots \\
 & + \lambda_m(R_m - d_{m1}w_1 - d_{m2}w_2 - \dots - d_{mp}w_p) \\
 & + \lambda(1 - \sum_{i=1}^p w_i).
 \end{aligned} \tag{15}$$

Taking derivatives of Z' with respect to w_1, \dots, w_p , $\lambda_1, \dots, \lambda_m$ and λ and setting them equal to zero will yield the following set of simultaneous equations (where $E(R_i) = E_i = \mu_i$ is used for the expected return μ_i). If $w_i < 0$ then we can again begin by striking out corresponding rows or columns and start the process again. This may now become very cumbersome and it is probably best to use the QP algorithm which can handle any amount of inequality constraints, equality constraints (in the same way as above) and upper and lower bounds of the type

$$L_i \leq w_i \leq U_i, i = 1, \dots, p.$$

The LP algorithm developed by Sharpe has been a favourite for many years because of the availability of LP (Linear Programming) routines. The Sharpe algorithm is as follows.

$$Max Z = \phi\mu_p - \sigma_p^2 \tag{16}$$

$$= \phi\mathbf{W}'\boldsymbol{\mu} - \mathbf{W}'\boldsymbol{\Sigma}\mathbf{W} \tag{17}$$

subject to

$$(a) \sum_{i=1}^p w_i = 1.$$

(b) Any equality constraints of the type given in (14).

(c) Any other inequality constraints (normal for any optimization routine)

(d) Bounds of the type $L_i \leq w_i \leq U_i$, $i = 1, \dots, p$.

But this seems exactly like a QP algorithm with ϕ varying from zero to infinity.

Sharpe, however starts at infinity. Since, in general, returns and variances are small quantities (a monthly return of, say 2%, is large) by choosing a large value of ϕ say a number like 1000 then the term $\phi\mu_p$ will dominate

$$Z = \phi\mu_p - \sigma_p^2$$

and we can ignore the term σ_p^2 . Thus Sharpe's formulation is

$$\begin{aligned} \text{Max } Z &= \phi\mu_p \\ &= \phi\mathbf{W}'\boldsymbol{\mu} \end{aligned} \tag{18}$$

$$= \phi \sum_{i=1}^p w_i \mu_i \tag{19}$$

subject to

$$(a) \sum_{i=1}^p w_i = 1.$$

(b) Any equality constraints of the type given in (14).

(c) Any other inequality constraints (normal for any optimization routine).

(d) Bounds of the type $L_i \leq w_i \leq U_i$, $i = 1, \dots, p$.

This is now a straight forward LP –Linear Programming problem which is routinely available on mainframe computers and PC's. Once a starting point has been found Sharpe

developed an algorithm using calculus to move down the Efficient Frontier generating all Corner Portfolios. A corner portfolio is one where a stock either enters or leaves the portfolio. These will be specific values of ϕ down the Efficient Frontier. For any other values of ϕ the formulas given by (11) and (13) can be used to compute the respective mean μ_p and variance σ_p^2 . The Corner Portfolio algorithm by Sharpe is available in the Jagger Library in his book *Portfolio Theory and Capital Markets*. For his contribution to Portfolio Theory Sharpe was also awarded the Nobel Prize in Economics and Finance.

When the constraints are of the type $(0, 1)$, that is $0 \leq w_i \leq 1$, then the first point in the Efficient Frontier (the one on the extreme right) is that stock with the largest return. Note the returns here are daily returns (log returns). In the computer program $X(i)$ is used in place of w_i , that is

$$w_i = X(i) \text{ and } N = p.$$

Notice that some of the returns are negative and others are close to zero. Should one really consider any of these stocks? But, remember the variances and covariances also play an important role in determining the efficient frontier.

7.3 The Capital Market Line

The Capital Market Line (CML) gives a satisfactory solution of how to use the Efficient Frontier once it has been generated. Suppose it is possible to borrow or lend any amount of money at the fixed interest rate R_f . The CML follows by drawing a straight line out from the riskless rate R_f into the (μ_p, σ_p^2) space. This line is then swung (upwards or

downwards) until it is just tangent to the Efficient Frontier. This yields the point M in Figure 3. Clearly if R_f or the Efficient Frontier moves then M will move accordingly.

Points between R_f and M represent lending portfolios (you are lending money to the bank). The portfolios are comprised by varying proportions of R_f and M . A point like H , located halfway between R_f and M , represents a portfolio of half R_f and half M .

Points on the CML line, that lie above M , represent borrowing portfolios. Their creation requires borrowing money at rate R_f to increase total investable capital. The total investable capital is then invested in M , which means that both return and risk (variance) is increased along the CML. (Note that the borrowing rate is often much higher than the lending rate R_f - such different rates can be accommodated).

The point M can easily be found by calculus and was derived by Lintner. If θ is the slope of the line from R_f to any stock in the Efficient Frontier as is shown in Figure 3 then maximizing the slope θ will give the point M . where M must therefore be on the boundary of the Efficient Frontier. The portfolio derived using the stocks that make up point M is called the Optimum Portfolio. For his contribution to Portfolio Theory deriving the Optimum Portfolio Lintner was awarded the Nobel Prize for Economics and Finance.

7.3.1 Estimation of μ and Σ .

We are dealing with Time Series data. Ordinary sampling procedures will not help us. We also know that both μ and Σ changes over time. Finance researchers in Europe believe a time period of 4 to 5 years is the maximum time period in which one can consider stable estimates. They are however vague about the starting point of the five years or whether

we are already in a new cycle. They, therefore prefer to use some smoothing techniques. Finance researches in the USA are now using Bayesian Estimation procedures using prior distributions of past data to compute posterior distributions of present and/or future data. It is, however common belief that daily data should be used. This area of research is at present under investigation for stocks on the JSE by the Econometrics and Finance Unit, of the Department of Statistical Sciences, at UCT.

7.4 The Index Model of Sharpe

Let the return (log) of particular stock be

$$R_t = \log P_t - \log P_{t-1}, \quad t = 1, \dots, N \quad (1)$$

where the time t is sufficiently large enough for R_t to follow a Normal $N(\mu_r, \sigma_r^2)$ distribution. Similarly let

$$X_t = I_t = \log I_t - \log I_{t-1}, \quad t = 1, \dots, N \quad (2)$$

be the return of the market proxy (usually the JSE overall index - JSE-OVER - or any of the other sector indices) be such that it also follows a Normal distribution $N(\mu_I, \sigma_I^2)$ distribution. We then have that

$$\begin{pmatrix} R_t \\ I_t \end{pmatrix} \sim N \left[\begin{pmatrix} \mu_r \\ \mu_I \end{pmatrix}, \begin{pmatrix} \sigma_r^2 & \sigma_{rI} \\ \sigma_{Ir} & \sigma_I^2 \end{pmatrix} \right] \quad (3)$$

has a bivariate normal distribution. Where $\sigma_{rI} = \sigma_{Ir}$ is the covariance between the return of stock r and the market proxy I . From the properties of the bivariate normal distribution

$$E(R_t/I_t) = \alpha + \beta I_t \quad (4)$$

and

$$\begin{aligned} \text{var}(R_t/I_t) &= \sigma_{r,I}^2 = \sigma_I^2(1 - \rho^2) \\ &= \sigma^2. \end{aligned} \quad (5)$$

where ρ is the correlation between the stock return and the market proxy. The model can thus be written as

$$R_t = \alpha + \beta I_t + e_t \quad (6)$$

where

$$e_t \sim N(0, \sigma^2). \quad (7)$$

It is usually assumed that the disturbances e_t are independently distributed over time, that is

$$E(e_t e_s) = 0, \text{ for } t \neq s. \quad (8)$$

The parameter β plays an important role in the area of finance. It can be used as a measure of the volatility of the security, relative to the market proxy. If β is greater than *one*, then, when the market rises, it is obvious that the return of a security will rise more rapidly than the return on the market. Similarly, if the market falls, the return on the security will fall more rapidly than the return on the market, and thus the security can be regarded as more volatile and hence more risky than the market. The converse is true if β is less than *one*. Thus β (or beta) is called the systematic risk of a security and has been used extensively in practice and in the literature for constructing and analyzing market portfolios.

The model (4) has received much discussion in the recent literature. We illustrate some important aspects of the assumptions of the model given in (4) which implies (3),

often called the market model, that is

$$E(R_t/I_t) = \alpha + \beta I_t.$$

- (i) The bivariate normal assumption (or multivariate normal for several securities and/ or indices) appears to be well accepted in the literature.
- (ii) If the bivariate normality is not feasible then the linearity assumption (4) appears to be very well satisfied (Fama, Fisher, Jensen and Roll).
- (iii) The assumption is made that the beta coefficient is fairly stable over time, especially as the length of the period under consideration increases .This implies that historical data can be used to estimate the parameters. Is this assumption reasonable? Suppose it is not, then can we do something to model this stochastic behavior of beta?
- (iv) Even if the beta coefficients do change over time, the ranking of the securities in order of risk does not. This is most relevant if the index model is used for portfolio optimization.
- (v) The beta coefficients do give a fairly good measure of the risk inherent in a security.
- (vi) The value of beta in any period can be related to some fundamental characteristics of the firm in that period.

- (vii) The error terms e_t , if not normally distributed, then still $E(e_t) = 0$, $E(e_t)^2 = \sigma^2$ and $E(e_t e_s) = 0$, $t \neq s$ so that least squares estimates can be used.

There are many more assumptions and many have been tested. A vast literature exists. It is now well known that the distributions of daily stock market returns for some securities have longer tails than what can be expected from the normal distribution. The student's t -distribution has become a popular substitute (MacDonald and Newey (1988)).

We assume that we are dealing with monthly returns (log) so that the normal assumption is a realistic approximation of the real world. If this is not so then we will use the least squares assumption (vii) above and invoke the Gauss - Markoff Theorem. (Hossain (2005))

7.5 The Single Index Model of Sharpe.

We formally write the market model as

$$R_t = \alpha + \beta I_t + e_t, \quad t = 1, \dots, N \quad (1)$$

where the time periods t is the choice of the portfolio manager. The assumptions are

$$E(e_t^2) = \sigma_e^2 \quad (2)$$

$$E(e_t e_s) = 0, \quad t \neq s = 1, \dots, N.$$

This is our assumption for a single stock R_t . A portfolio is made up of several stocks - a large number indeed. Let the i^{th} stock be

$$R_{it} = \alpha_i + \beta_i I_t + e_{it}, \quad i = 1, \dots, p; \quad t = 1, \dots, N, \quad (3)$$

where all stocks are regressed on the same single index I (say the JSE-OVER). Our assumptions now become

$$E(e_{it}^2) = \sigma_{ei}^2 \quad (4)$$

$$E(e_{it}e_{is}) = 0, \quad t \neq s = 1, \dots, N, \quad (5)$$

$$E(e_{it}I_t) = 0, \quad t = 1, \dots, N, \quad (6)$$

$$E(e_{it}e_{jt}) = 0, \quad t = 1, \dots, N. \quad (7)$$

Equation (4) assumes each stock has its own variance for the error term. Equation (5) assumes that the error terms of each stock are independent over time the usual assumption of no serial (auto) correlation - no time series. Equation (6) assumes that the errors of each stock and the explanatory variable I are uncorrelated - this is the usual assumption in regression. Equation (7) assumes that the error terms of the stocks are uncorrelated, that is the stocks are only related through their mutual relationship with the index I . Let

$$E(I) = \mu_I \quad \text{and} \quad \text{var}(I) = \sigma_I^2 \quad (8)$$

be the mean and variance of the index. For each stock let

$$E_i = E(R_i) = \alpha_i + \beta_i \mu_I. \quad (9)$$

The variance is

$$\begin{aligned} \text{var}(R_i) &= \text{var}(\alpha_i + \beta_i I + e_i) \\ &= \beta_i^2 \sigma_I^2 + \sigma_{ei}^2 \\ &= \sigma_{ii} = \sigma_i^2 \end{aligned} \quad (10)$$

and the covariance is

$$\begin{aligned}
 cov(R_i R_j) &= E\{(R_i - E(R_i))(R_j - E(R_j))\} \\
 &= E\{\beta_i(I - \mu_I) + e_i\}\{\beta_j(I - \mu_I) + e_j\} \\
 &= \beta_i \beta_j \sigma_I^2
 \end{aligned} \tag{11}$$

Our standard portfolio problem then becomes

$$\begin{aligned}
 Max Z &= \phi \mu_p - \sigma_p^2 \\
 &= \phi \mathbf{W}' \boldsymbol{\mu} - \mathbf{W}' \boldsymbol{\Sigma} \mathbf{W} \text{ subject to} \\
 \sum_{i=1}^p w_i &= 1.
 \end{aligned} \tag{12}$$

where

$$\begin{aligned}
 \mu_p &= \sum_{i=1}^p w_i E_i \\
 &= \sum_{i=1}^p (\alpha_i + \beta_i \mu_I)
 \end{aligned} \tag{13}$$

and

$$\begin{aligned}
 \sigma_p^2 &= \mathbf{W}' \boldsymbol{\Sigma} \mathbf{W} \\
 &= \sum_{i=1}^p \sum_{j=1}^p w_i w_j \sigma_{ij}
 \end{aligned}$$

with

$$\begin{aligned}
 \sigma_{ii} &= \sigma_i^2 \\
 &= \beta_i^2 \sigma_I^2 + \sigma_{ei}^2 \text{ and} \\
 \sigma_{ij} &= \beta_i \beta_j \sigma_I^2.
 \end{aligned} \tag{14}$$

Thus to calculate the efficient frontier we need to estimate

$$(1) \alpha_i, \beta_i, \sigma_{ei}^2, \quad i = 1, \dots, p$$

and

$$(2) \mu_I \quad \text{and} \quad \sigma_I^2.$$

The total number of quantities to estimate is

$$3p + 2.$$

It is important to remember that the total number of quantities to estimate for the normal portfolio problem is

$$\frac{1}{2}p(p + 1).$$

For $p = 19$ we have $3(19) + 2 = 59$ compared to $19(19 + 1)/2 = 190$. For the normal portfolio problem we also need to estimate $E(R_i) = \mu_i$, while for the Sharpe Index model we need to estimate

$$E_i = \alpha_i + \beta_i \mu_I. \quad (15)$$

There are considerable advantages why we should use (15). We could use different indices. We could use overseas favourites like the Dow-Jones, the Financial Times, the Dax or the Nikkei. More important by simply changing μ_I or σ_I^2 , or both, we can simulate different portfolios. We can use regression or time series techniques to forecast $E(I) = \mu_I$ and set confidence intervals on such forecasts.

7.6 Estimation of the parameters using Least Squares.

We will deal only with the multi-index model since the single index model is a special case.

The multi-index model can be written as

$$R_{it} = \alpha_i + \beta_{i1}I_{1t} + \beta_{i2}I_{2t} + \dots + \beta_{iM}I_{Mt} + e_{it}, \quad (1)$$

$$i = 1, \dots, p, \quad t = 1, \dots, N,$$

with the following assumptions

$$E(e_{it}^2) = \sigma_{ei}^2 \quad (2)$$

$$E(e_{it}e_{is}) = 0, \quad t \neq s = 1, \dots, N, \quad (3)$$

$$E(e_{it}I_{jt}) = 0, \quad j = 1, \dots, M, \quad t = 1, \dots, N, \quad (4)$$

$$E(e_{it}e_{jt}) = \sigma_{ij}, \quad t = 1, \dots, N. \quad (5)$$

$$E(I_{jt}I_{kt}) = c_{jk}, \quad j, k = 1, \dots, M. \quad (6)$$

We now formulate the problem in standard regression notation. Let

$$\mathbf{Y}_i = \begin{pmatrix} R_{1i} \\ \vdots \\ R_{Ni} \end{pmatrix}, \quad i = 1, \dots, p \quad (7)$$

$$\mathbf{X} = \begin{pmatrix} 1 & I_{11} & \dots & I_{1M} \\ \vdots & \vdots & \vdots & \vdots \\ 1 & I_{N1} & \dots & I_{NM} \end{pmatrix} \quad (8)$$

and

$$\boldsymbol{\beta}_i = \begin{pmatrix} \beta_{1i} \\ \vdots \\ \beta_{Mi} \end{pmatrix}, \quad \mathbf{e}_i = \begin{pmatrix} e_{1i} \\ \vdots \\ e_{Ni} \end{pmatrix}, \quad (9)$$

where we note carefully the switch of the subscripts $i = 1, \dots, p$ and $t = 1, \dots, N$. Then

$$\mathbf{Y}_i = \mathbf{X}\boldsymbol{\beta}_i + \mathbf{e}_i, \quad i = 1, \dots, p \quad (10)$$

is in the form of standard least squares problem. Each equation can be solved individually to obtain the estimates

$$\begin{aligned}\hat{\beta}_i &= (\mathbf{X}'\mathbf{X})^{-1}\mathbf{X}'\mathbf{Y}_i, i = 1, \dots, p \\ \hat{e}_i &= \mathbf{Y}_i - \mathbf{X}\hat{\beta}_i, i = 1, \dots, p.\end{aligned}\tag{11}$$

We can write all the p equations as a single equation as follows. Let

$$\mathbf{Y} = (\mathbf{Y}_1 \dots \mathbf{Y}_p)\tag{12}$$

$$\boldsymbol{\beta} = (\boldsymbol{\beta}_1 \dots \boldsymbol{\beta}_p), \text{ and}$$

$$\mathbf{e} = (\mathbf{e}_1 \dots \mathbf{e}_p)$$

then

$$\mathbf{Y} = \mathbf{X}\boldsymbol{\beta} + \mathbf{e}.$$

The normal equations can be written

$$\mathbf{X}'\mathbf{X}\boldsymbol{\beta} = \mathbf{X}'\mathbf{Y}$$

and

$$\hat{\boldsymbol{\beta}} = (\mathbf{X}'\mathbf{X})^{-1}\mathbf{X}'\mathbf{Y}$$

with

$$\begin{aligned}\hat{\mathbf{e}} &= \mathbf{Y} - \mathbf{X}\hat{\boldsymbol{\beta}} \\ &= \begin{pmatrix} \hat{e}_{11} & \dots & \hat{e}_{1p} \\ \vdots & \vdots & \vdots \\ \hat{e}_{N1} & \dots & \hat{e}_{Np} \end{pmatrix}\end{aligned}$$

and is an $(N \times p)$ residual matrix. The $(p \times p)$ moment matrix is then

$$\hat{\mathbf{e}}'\hat{\mathbf{e}}.$$

and the sample covariance matrix is

$$S = \frac{1}{N - M - 1} \hat{\mathbf{e}}' \hat{\mathbf{e}}.$$

Some computer packages like TSP allows for the above simultaneous relationship. The covariance matrix of the i^{th} sample estimate $\hat{\beta}_i$ is

$$E(\hat{\beta}_i - \beta_i)(\hat{\beta}_i - \beta_i)' = \sigma_{ei}^2 (\mathbf{X}'\mathbf{X})^{-1}$$

and the covariance matrix between the i^{th} and j^{th} sample estimates $\hat{\beta}_i$ and $\hat{\beta}_j$ is

$$E(\hat{\beta}_i - \beta_i)(\hat{\beta}_j - \beta_j)' = \sigma_{eij} (\mathbf{X}'\mathbf{X})^{-1}.$$

The covariance matrix of $\hat{\beta} = (\hat{\beta}_1 \dots \hat{\beta}_p)$ is

$$E(\hat{\beta} - \beta)(\hat{\beta} - \beta)' = \Omega \otimes (\mathbf{X}'\mathbf{X})^{-1}.$$

where $E(\mathbf{e}'\mathbf{e}) = \Omega$. The whole system of equations

$$\mathbf{Y}_i = \mathbf{X}\beta_i + \mathbf{e}_i, \quad i = 1, \dots, p, \quad (13)$$

$$E(\mathbf{e}_i) = \mathbf{0}, \quad i = 1, \dots, p,$$

$$E(\mathbf{e}_i \mathbf{e}_j) = \sigma_{ij} \mathbf{I} \quad i, j = 1, \dots, p,$$

can also be written as

$$\begin{pmatrix} \mathbf{Y}_1 \\ \mathbf{Y}_2 \\ \vdots \\ \mathbf{Y}_p \end{pmatrix} = \begin{pmatrix} \mathbf{X} & \mathbf{0} & \dots & \mathbf{0} \\ \mathbf{0} & \mathbf{X} & \vdots & \vdots \\ \vdots & \vdots & \vdots & \vdots \\ \mathbf{0} & \mathbf{0} & \dots & \mathbf{X} \end{pmatrix} \begin{pmatrix} \beta_1 \\ \beta_2 \\ \vdots \\ \beta_p \end{pmatrix} = \begin{pmatrix} \mathbf{e}_1 \\ \mathbf{e}_2 \\ \vdots \\ \mathbf{e}_p \end{pmatrix} \quad (14)$$

or more compactly as

$$\mathbf{Y}^* = \mathbf{X}^* \beta^* + \mathbf{e}^*. \quad (15)$$

Note however that under this formulation

$$E(\mathbf{e}^* \mathbf{e}^{*\prime}) = \Omega \otimes \mathbf{I} = \Psi \quad (16)$$

so that to estimate β^* we use the Aitken estimator

$$\hat{\beta}^* = (\mathbf{X}^{*\prime} \Psi^{-1} \mathbf{X}^*)^{-1} \mathbf{X}^{*\prime} \Psi^{-1} \mathbf{Y}^* \quad (17)$$

with covariance matrix

$$\text{cov}(\hat{\beta}^*) = (\mathbf{X}^* \Psi^{-1} \mathbf{X}^*)^{-1}. \quad (18)$$

Since the \mathbf{X} matrix is the same for all the relationships i it can be shown (see Goldberger page 247) that

$$\hat{\beta}_i^* = \hat{\beta}_i = (\mathbf{X}'\mathbf{X})^{-1} \mathbf{X}'\mathbf{Y}_i \quad (19)$$

which is the equation by equation estimate of β_i . Also

$$(\mathbf{X}^* \Psi^{-1} \mathbf{X}^*)^{-1} = \Omega \otimes (\mathbf{X}'\mathbf{X})^{-1}. \quad (20)$$

7.7 Empirical Study: Single Index

7.7.1 The Data

For this empirical analysis, the data set used was the SIDS batch.

7.7.2 Study Objectives

In this study we construct efficient frontiers for the 4 modelling methods applied on our portfolio of stocks and then conduct relative comparisons. We empirically show that the mean variance frontier using $GARCH(1, 1)$ inputs is positioned to the left of the OLS , Kalman Filter and $ARMA$ (Autoregressive Moving Average) efficient frontiers in the Single Index case. This is in support of the central hypothesis regarding $GARCH(1, 1)$'s risk-return superiority. The optimisation programs used to construct the efficient frontiers have been written in the programming language Fortran by Professor Troskie of the Statistical Sciences Department at the University of Cape Town.

7.7.3 Methodology

All the inputs required for risk-return construction are obtained from previous 4 empirical studies regarding the various modelling chapters on our portfolio of 9 stocks. The inputs include the mean of the log returns for the 9 shares. The variance of regressions are extracted from the portfolio of stocks for each modelling technique. For our case it's four methods. Similarly we collect the beta coefficients concerning the explanatory variable JSE for each

of the 9 stock models within each estimation technique. Finally we also employ the variance of the index, JSE as an input into our optimisation program.

The optimisation algorithm is then carried out on each set of results for each estimation technique, i.e. in our case this will be 4 iterations. We then gather and save the mean and variance output in Microsoft excel and plot the 4 different sets of means vs variances together.

7.7.4 Primary Findings

The above methodology was implemented for the *ARMA*, *OLS*, State Space and *GARCH*(1, 1) models. The following efficient frontiers were constructed for the 4 respective models in discussion, below in figure 7.1

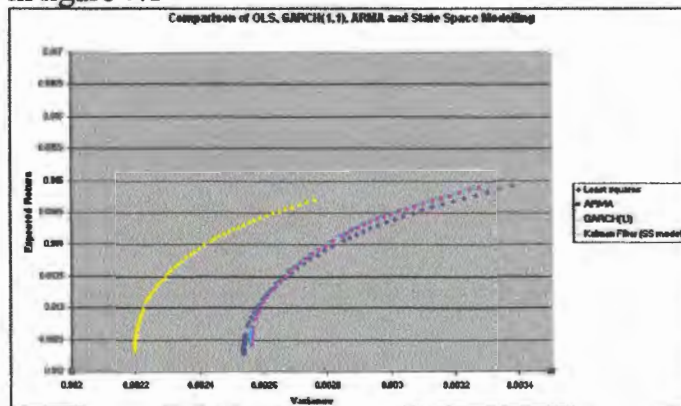


Figure 7.1: Single index Efficient Frontiers

Figure 7.1 above illustrates a shift of the *GARCH*(1, 1) efficient frontier to the left, which is consistent with our central hypothesis. The lower variance is largely a consequence of the lower error variances from the regression, which is used in the computation of the variance of the individual securities. Thus it is clear from above that for every level of expected return, the *GARCH*(1, 1) frontier exhibits the most minimum level of risk compared to

the other frontiers. In a later study we will show that this is empirically consistent in the multiple index framework as well.

From figure 7.1 the frontiers for the Kalman filter, ARMA and least squares coincide at the point (0.00263617; 0.01337822). To the left of this point, least squares exhibits the superior results of the 3 currently in mention, whilst the ARMA is the worst. To the right, the converse holds true, where the least squares is seen as the weakest estimator of risk, whilst it is the Kalman filter in this instance that is the superior of the 3. Moreover, the frontier using Kalman Filter estimation attains larger returns as indicated by its length compared with the frontier generated with the least squares and ARMA inputs.

7.7.5 Conclusion

In this study it is concluded that the left-shifting role of $GARCH(1, 1)$ (General Autoregressive Conditional Heteroskedastic) model's efficient frontier provides a graphic device for optimizing portfolio allocations. It was shown that within the single index framework there occurred apparent shifts to the left in the efficient frontier. Thus for every level of expected return, the $GARCH(1, 1)$ estimates exhibit the most minimum level of risk compared to the other frontiers in the context of the South African market. In conclusion, the $GARCH(1, 1)$ estimation inputs result in the most efficient single index mean-variance frontiers for financial portfolios in the South African context. (Hossain (2005))

7.8 GARCH(1,1) versus GARCH Extensions

7.8.1 The Data

For this empirical analysis, the data set used was the SIDS batch.

7.8.2 Study Objectives

Having established $GRACH(1,1)$'s risk-return superiority in the Sharpe Single Index framework amongst rival techniques, we go onto explore the consistency of this relationship within a $GARCH$ context in the Sharpe Single Index setting. Hence in this study we construct efficient frontiers for the $GARCH(1,1)$ modelling extensions on our portfolio of 9 stocks. The aim is to investigate the superiority of $GARCH(1,1)$ versus its extensions, namely $EGARCH(1,1)$, $PARCH(1,1)$, $TARCH(1,1)$ and $C - GARCH(1,1)$. Hence we attempt to demonstrate the supremacy of $GARCH(1,1)$ with regards to risk (variance) within the GARCH modelling paradigm. This will be explored in the Sharpe single index setting. The optimisation programs used to construct the efficient frontiers have been written in the programming language Fortran by Professor Troskie of the Statistical Sciences Department at the University of Cape Town.

7.8.3 Methodology

First we gather the β^i s of the JSE index for all 9 responses for each of the 5 $GARCH$ techniques. Next we extract the SE's of regression of our portfolio of stocks for each of the 5 $GARCH$ estimation techniques. The variance of JSE and the means of the 9 shares remain the same throughout each mean-variance construction phase.

We then run the optimisation program on each set of results for each *GARCH* technique, i.e. in our case will be 5 times. We then gather and save the mean and variance output in Microsoft excel and plot the four different means vs variance together.

7.8.4 Primary Findings

The above methodology was implemented for the *EGARCH*(1, 1), *TARCH*(1, 1), *PARCH*(1, 1), *C – GARCH*(1, 1) and the *GARCH*(1, 1) models. The following efficient frontiers were constructed for the 5 respective models in discussion, below in figure 7.2

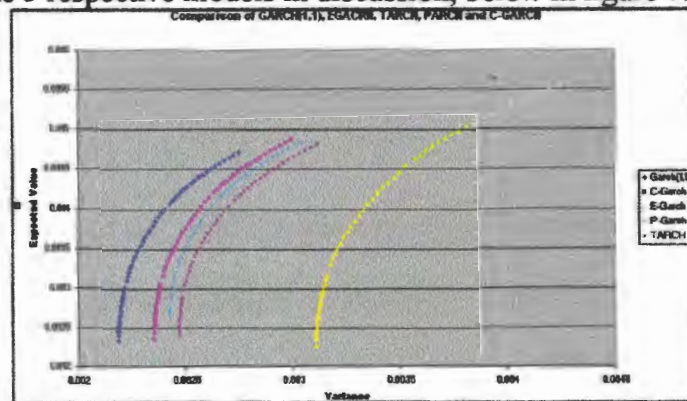


Figure 7.2: GARCH(1,1) vs Extensions - Single Index

Figure 7.2 above illustrates a shift of the efficient frontier using *GARCH*(1, 1) estimates to the left. Once again, this is consistent with the hypothesis set out at the start of this thesis regarding *GARCH*(1, 1)'s superiority over its extensions. The lower variance (view of risk) of *GARCH*(1, 1) is largely a consequence of the lower error variances of regressions computed for the portfolio of 9 stocks. Figure 7.2 also demonstrates that the *E – GARCH*(1, 1) model exhibits the worst model of risk of the 5 models presented above. This can be attributed to the large standard errors produced by its models and the less significant β estimates.

7.8.5 Conclusion

With regards to this study, it has been further established that the left-shifting role of the $GARCH(1, 1)$ model's mean-variance frontier provides a graphic device for optimizing portfolio allocations. This time the investigation was conducted within the $ARCH/GARCH$ framework where the $GARCH(1, 1)$ model was compared to its extensions namely, $E-GARCH(1, 1)$, $TARCH(1, 1)$, $C-GARCH(1, 1)$ and finally $PARCH(1, 1)$. It was empirically shown that the $GARCH(1, 1)$ model estimates exhibit the most minimum level of risk compared to the risk-return structures of its extensions within a single index setting. As a result it is established that the $GARCH(1, 1)$ model exhibits the most superior mean-variance frontiers when compared with its similar extensions within the single index framework (Hossain (2005)).

7.9 Generalisation of the Markowitz Formulations

The classical Markowitz formulation of the portfolio problem is

$$\begin{aligned} \text{Min } \sigma_p^2 &= \mathbf{W}'\Sigma\mathbf{W} \\ &= \sum_{i,j=1}^p w_i w_j \sigma_{ij} \end{aligned} \quad (1)$$

subject to

$$\begin{aligned} \mu_p &= \mathbf{W}'\boldsymbol{\mu}_p \\ &= \sum_{i=1}^p w_i \mu_i \end{aligned} \quad (2)$$

$$= E_k \quad (3)$$

and subject to

$$\sum_{i=1}^p w_i = 1 \quad (4)$$

and

$$0 \leq w_i \leq 1 \quad (5)$$

where

$$\Sigma = (\sigma_{ij}), \quad i, j = 1, \dots, p \quad (6)$$

Short sales (negative w_i) are common in markets around the world and it is no longer necessary to restrict portfolios to be legitimate. The generalised formulation of the Markowitz model is

$$\begin{aligned} \text{Min } \sigma_p^2 &= \mathbf{W}'\Sigma\mathbf{W} \\ &= \sum_{i,j=1}^p w_i w_j \sigma_{ij} \end{aligned} \quad (7)$$

subject to

$$\begin{aligned} \mu_p &= \mathbf{W}'\boldsymbol{\mu}_p \\ &= \sum_{i=1}^p w_i \mu_i \end{aligned} \quad (8)$$

$$= E_k \quad (9)$$

and subject to

$$\sum_{i=1}^p w_i = 1 \quad (10)$$

where

$$\Sigma = \begin{pmatrix} \sigma_{11} & \sigma_{12} & \dots & \sigma_{1p} \\ \sigma_{21} & \sigma_{22} & \dots & \sigma_{2p} \\ \vdots & \vdots & \ddots & \vdots \\ \sigma_{p1} & \sigma_{p2} & \dots & \sigma_{pp} \end{pmatrix} \quad (11)$$

where $\sigma_{ii} = \sigma_i^2$ is the variance of the i 'th stock and σ_{ij} is the covariance between the i 'th and j 'th stock. In this formulation no restrictions are put on the weights ω_i . Since the weights can take negative values they include the possibility of short selling.

7.10 The Multiple Index Model of Sharpe.

The multi - index model can be written as

$$R_{it} = \alpha_i + \beta_{i1}I_1 + \beta_{i2}I_2 + \dots + \beta_{iM}I_M + e_{it}, \quad (1)$$

$$i = 1, \dots, p, \quad t = 1, \dots, N,$$

with the following assumptions

$$E(e_{it}^2) = \sigma_{ei}^2 \quad (2)$$

$$E(e_{it}e_{is}) = 0, \quad t \neq s = 1, \dots, N, \quad (3)$$

$$E(e_{it}I_{jt}) = 0, \quad j = 1, \dots, M, \quad t = 1, \dots, N, \quad (4)$$

$$E(e_{it}e_{jt}) = 0, \quad t = 1, \dots, N. \quad (5)$$

$$E(I_{jt}I_{kt}) = c_{jk}, \quad j, k = 1, \dots, M \quad (6)$$

These assumptions are identical to the Single Index Model where in equation (4) we now also assume that the disturbance term e_{it} is also independent of the Indices I_j , $j = 1, \dots, M$. This again is a normal assumption in regression. We further assume that the Indices are dependent with covariances given by c_{jk} (see equation (6)).

Let

$$E_i = E(R_i) = \alpha_i + \beta_{i1}I_1 + \dots + \beta_{iM}I_M, \quad (7)$$

$$i = 1, \dots, p.$$

For portfolio $P = W'R = \sum_{i=1}^p w_i R_i$ we have that

$$\mu_p = \sum_{i=1}^p w_i E_i \quad (8)$$

and

$$\sigma_p^2 = \sum_{i=1}^p w_i w_j \sigma_{ij} \quad (9)$$

$$= \sum_{i=1}^p w_i^2 \sigma_{ei}^2 \quad (10)$$

$$+ \sum_{i=1}^p w_i^2 \sum_{k,l} \beta_{ik} \beta_{il} c_{kl} \\ + \sum_{i \neq j} \sum_j w_i w_j \sum_{k,l} \beta_{ik} \beta_{jl} c_{kl} \\ = \sum_{i=1}^p w_i^2 \sigma_{ei}^2 \quad (11)$$

$$+ \sum_i \sum_j w_i w_j \sum_k \sum_l \beta_{ik} \beta_{jl} c_{kl}$$

$$= \sum_{i=1}^p w_i^2 \sigma_{ei}^2 \quad (12)$$

$$+ \sum_k \sum_l \sum_i \sum_j w_i \beta_{ik} w_j \beta_{jl} c_{kl} \quad (13)$$

$$= \sum_{i=1}^p w_i^2 \sigma_{ei}^2 + \quad (14)$$

$$+ \sum_k \sum_l \left(\sum_i w_i \beta_{ik} \right) \left(\sum_j w_j \beta_{jl} \right) c_{kl}$$

$$= \sum_{i=1}^p w_i^2 \sigma_{ei}^2 + \sum_k \sum_l \beta_{pk} \beta_{pl} c_{kl}$$

where

$$\begin{aligned}\beta_{pk} &= \sum_{i=1}^p w_i \beta_{ik}, k = 1, \dots, M \\ \beta_{pl} &= \sum_{j=1}^p w_j \beta_{jl}, l = 1, \dots, M.\end{aligned}$$

Note that these two are the same and need only to be computed once. The Sharpe formulation is then (minimize instead of maximize)

$$\text{Min } Z = -\Lambda \mu_p + \sigma_p^2$$

subject to

$$\begin{aligned}\beta_{p1} &= \sum_{i=1}^p w_i \beta_{i1} \\ \beta_{p2} &= \sum_{i=1}^p w_i \beta_{i2} \\ &\vdots \\ \beta_{pM} &= \sum_{i=1}^p w_i \beta_{iM} \\ \sum_{i=1}^p w_i &= 1,\end{aligned}$$

and any other equality, in-equality constraints or bounds. Ignoring for the moment any other constraints and bounds the objective function becomes

$$\begin{aligned}
 \text{Min } Z' &= -\Lambda\mu_p + \sigma_p^2 \\
 &+ \lambda_1(\beta_{p1} - \sum_{i=1}^p w_i\beta_{i1}) \\
 &+ \lambda_2(\beta_{p2} - \sum_{i=1}^p w_i\beta_{i2}) \\
 &\vdots \\
 &+ \lambda_M(\beta_{pM} - \sum_{i=1}^p w_i\beta_{iM}) \\
 &+ \lambda_f(1 - \sum_{i=1}^p w_i).
 \end{aligned}$$

The solution requires that the partial derivative of Z' with respect to each variable be set to zero.

For each i from 1 to p ,

$$\frac{\partial Z'}{\partial w_i} = 2\sigma_{ei}^2 - E_i\Lambda - \lambda_1\beta_{i1} - \dots - \lambda_M\beta_{iM} = 0.$$

For each j from 1 to M ,

$$\frac{\partial Z'}{\partial \beta_{pj}} = 2\beta_{p1}c_{j1} + 2\beta_{p2}c_{j2} + \dots + 2\beta_{pM}c_{jM} + \lambda_j = 0.$$

For each j from 1 to M ,

$$\frac{\partial Z'}{\partial \lambda_j} = \beta_{pj} - \beta_{1j}w_1 - \dots - \beta_{pj}w_p = 0.$$

For λ_f ,

$$\frac{\partial Z'}{\partial \lambda_f} = 1 - w_1 - \dots - w_p = 0.$$

The system of linear equations that need to be solved is

$$AX = B,$$

(where the top row is X and the entries not filled in are zeroes) with solution

$$X = A^{-1}B.$$

provided that no other equality, in-equality or bounds are imposed. If there are such restrictions they must be added to the systems of equations. The problem may then no longer be a calculus problem but rather a QP or a LP problem.

Sharpe has again given an algorithm for a LP solution by setting Λ to be a large number, and then computing the efficient frontier by varying Λ . Affleck-Graves wrote a computer program for the main-frame computer to implement this algorithm. This computer program is, however very sensitive to singular or near- singular values of the matrix A . Troskie has modified the program using a SVD to run on a PC. Troskie has also written a QP program for a PC. The listing below is given by the PC program written by Troskie using Sharpe's algorithm. The bounds in this case are

$$0 \leq w_i \leq 1$$

so that the first stock that enters the efficient frontier is the one with the largest return E_i .

We again use the weights $w_i = X_{(i)}$.

7.11 Empirical Study: Multiple Index

7.11.1 The Data

For this empirical analysis, the data set used was the MIDS batch.

7.11.2 Study Objectives

It has been established thus far that $GARCH(1, 1)$ is the superior technique in risk frontier construction amongst rival modelling techniques and within its own domain in a single index context. We now shift our focus to explore this relationship within the Sharpe multiple index framework. We now attempt to show that the mean variance frontier using Multiple Index (MI)- $GARCH$ inputs is positioned to the left of the OLS , Kalman Filter and $ARMA$ (Autoregressive Moving Average) efficient frontiers in the multiple index case, which is again in line with the central hypothesis. The optimisation programs used to construct the efficient frontiers have been written in the programming language Fortran by Professor Troskie of the Statistical Sciences Department at the University of Cape Town.

7.11.3 Methodology

Yet again we gather all the inputs relevant for the risk-return frontier construction, namely the β 's and the SE-of regressions for each of the 4 modelling techniques on our portfolio of stocks within the multiple index context. These inputs have already been calculated in the previous chapters regarding the respective modelling techniques. The results include the mean of the log returns for the 9 shares. Similarly we collect the β coefficients of the

3 indices namely, the JSE All share, Paladium and Implats for each of the 9 stocks within each estimation technique. Finally we also use the covariance of the 3 indices, as an input into our Fortran optimisation program.

The optimisation program is then applied on each set of results for each estimation technique, i.e. in our case this will be 4 times. We then gather and save the mean and variance output in Microsoft excel and plot the 4 different sets of means and variances against each other.

7.11.4 Primary Findings

Implementing the above methodology for the multiple index *ARMA*, regression, State Space and *GARCH*(1,1) models; the following efficient frontiers were constructed for the 4 respective models in discussion, below in figure 7.3

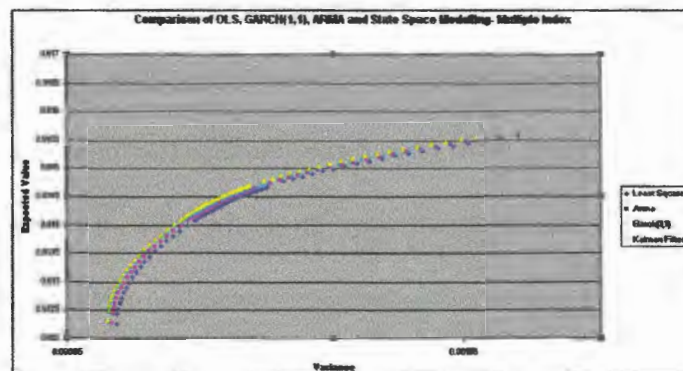


Figure 7.3: Multiple Index Efficient Frontiers

Figure 7.3 above shows that the efficient frontier using *MI – GARCH*(1,1) estimates is located to the left of the other 3 using, *ARMA*, *OLS* and State Space inputs, which is

the same conclusion as for the single index case (as in figure 7.1) and hence provides even further support the central hypothesis of this thesis.

The underlying reason for the leftward shift can be attributed on similar grounds as described for the single-index case. The effect of lower error variances due to regression translates into a covariance structure that produces lower variance estimates in principle. Thus it is clear from above that at every value of expected return, the $MI - GARCH(1, 1)$ estimates exhibit the most minimum level of risk compared to the other frontiers.

7.11.5 Conclusion

In this section it is concluded that the left-shifting role of $GARCH(1, 1)$ (General Autoregressive Conditional Heteroskedastic) model and its extensions of the efficient frontier provides a graphic device for optimizing portfolio allocations. It was shown that within the multiple and single index framework there occurred apparent shifts to the left in the efficient frontiers. Thus at every value of expected return, the $GARCH(1, 1)$ estimates exhibit the most minimum level of risk compared to the other frontiers in the context of the South African market. In conclusion, the $GARCH(1, 1)$ estimation inputs result in the most efficient mean-variance frontiers for financial portfolios in the South African context.

Chapter 8

The Troskie Hossain Innovations: The Improved Sharpe

8.1 Innovations to the Sharpe Index Model

Use of Bayesian estimates in the Sharp single index model

The posterior mean of the predictive distribution of a future observation of the return of the Jse would be the ideal choice. If we call this posterior predictive mean I_f and if

$$\mathbf{R}_f = \begin{pmatrix} R_{1f} \\ R_{2f} \\ \vdots \\ R_{pf} \end{pmatrix}$$

are the future predictive returns of the p -stocks; then setting a prior distribution for

$$\begin{pmatrix} \alpha \\ \beta \end{pmatrix} = \begin{pmatrix} \alpha \\ \beta_1 \\ \beta_2 \\ \vdots \\ \beta_p \end{pmatrix}$$

we can derive the posterior distribution of

$$\begin{pmatrix} \alpha \\ \beta \end{pmatrix}$$

from which we can derive the posterior distribution of the predictive distribution of \mathbf{R}_f which will depend on the posterior predictive mean I_f . We can then compute the posterior mean and posterior covariance matrix of \mathbf{R}_f and use this as our input for portfolio analysis. The rest of our calculations will be exactly the same as given above. An alternative semi Bayesian approach would be to use the confidence and predictive intervals of future values

of each of the elements of

$$\mathbf{R}_f = \begin{pmatrix} R_{1f} \\ R_{2f} \\ \vdots \\ R_{pf} \end{pmatrix}$$

with I_f as the value for the Jse-over index.

8.2 The Improved Sharpe Single Index Model by Troskie and Hossain.

The Sharpe single index model is formulated as

$$R_{it} = \alpha_i + \beta_i I_t + e_{it}, \quad i = 1, \dots, p; \quad t = 1, \dots, N, \quad (1)$$

$$E(e_{it}^2) = \sigma_{ei}^2 \quad (2)$$

$$E(e_{it}e_{is}) = 0, \quad t \neq s = 1, \dots, N, \quad (3)$$

$$E(e_{it}I_t) = 0, \quad t = 1, \dots, N, \quad (4)$$

$$E(e_{it}e_{jt}) = 0, \quad t = 1, \dots, N. \quad (5)$$

In vector notation

$$\mathbf{R}_t = \boldsymbol{\alpha} + \beta I_t + \mathbf{e}_t \quad t = 1, \dots, N \quad (6)$$

where

$$\mathbf{R}_t = \begin{pmatrix} R_{1t} \\ \vdots \\ R_{pt} \end{pmatrix}, \quad \boldsymbol{\alpha} = \begin{pmatrix} \alpha_1 \\ \vdots \\ \alpha_p \end{pmatrix}, \quad \beta = \begin{pmatrix} \beta_1 \\ \vdots \\ \beta_p \end{pmatrix} \quad \text{and} \quad \mathbf{e}_t = \begin{pmatrix} e_{1t} \\ \vdots \\ e_{pt} \end{pmatrix} \quad (7)$$

so that (conveniently dropping the index t)

$$E(\mathbf{R}) = \boldsymbol{\alpha} + \beta \mu_I \quad (8)$$

and

$$\text{cov}(\mathbf{e}) = \begin{pmatrix} \sigma_{e1}^2 & 0 & \dots & 0 \\ 0 & \sigma_{e2}^2 & \dots & 0 \\ \vdots & \vdots & \ddots & \vdots \\ 0 & \dots & \dots & \sigma_{ep}^2 \end{pmatrix}. \quad (9)$$

This implies that

$$\text{cov}(\mathbf{R}) = \sigma_I^2 \beta \beta' + \begin{pmatrix} \sigma_{e1}^2 & 0 & \dots & 0 \\ 0 & \sigma_{e2}^2 & \dots & 0 \\ \vdots & \vdots & \ddots & \vdots \\ 0 & \dots & \dots & \sigma_{ep}^2 \end{pmatrix}. \quad (10)$$

Here we have assumed that all e_{it} , $i = 1, \dots, p$, $t = 1, \dots, N$ are independent. We now relax this assumption by assuming that the disturbances e_{it} of the different stocks are dependent (correlated). There is evidence that this is indeed the case on the JSE, but the correlations are not high.

Thus

$$\begin{aligned} E(e_{it}e_{jt}) &= \sigma_{ij}, \quad i \neq j \\ &= \sigma_{ei}^2 \quad i = j \\ &= \sigma_{ii}. \end{aligned}$$

or

$$E(\mathbf{ee}') = \Omega = \begin{pmatrix} \sigma_{e1}^2 & \sigma_{12} & \dots & \sigma_{1p} \\ \sigma_{21} & \sigma_{e2}^2 & \dots & \sigma_{2p} \\ \vdots & \vdots & \ddots & \vdots \\ \sigma_{p1} & \dots & \dots & \sigma_{ep}^2 \end{pmatrix} \quad (11)$$

with

$$\text{cov}(\mathbf{R}) = \sigma_I^2 \beta \beta' + \Omega = \Phi \quad (12)$$

For portfolio $P = W'R$ we have

$$E(P) = W'(\alpha + \beta\mu_I) = \mu_p \quad (13)$$

of each of the elements of

$$\mathbf{R}_f = \begin{pmatrix} R_{1f} \\ R_{2f} \\ \vdots \\ R_{pf} \end{pmatrix}$$

with I_f as the value for the Jse-over index.

8.2 The Improved Sharpe Single Index Model by Troskie and Hossain.

The Sharpe single index model is formulated as

$$R_{it} = \alpha_i + \beta_i I_t + e_{it}, \quad i = 1, \dots, p; \quad t = 1, \dots, N, \quad (1)$$

$$E(e_{it}^2) = \sigma_{ei}^2 \quad (2)$$

$$E(e_{it}e_{is}) = 0, \quad t \neq s = 1, \dots, N, \quad (3)$$

$$E(e_{it}I_t) = 0, \quad t = 1, \dots, N, \quad (4)$$

$$E(e_{it}e_{jt}) = 0, \quad t = 1, \dots, N. \quad (5)$$

In vector notation

$$\mathbf{R}_t = \boldsymbol{\alpha} + \beta I_t + \mathbf{e}_t \quad t = 1, \dots, N \quad (6)$$

where

$$\mathbf{R}_t = \begin{pmatrix} R_{1t} \\ \vdots \\ R_{pt} \end{pmatrix}, \quad \boldsymbol{\alpha} = \begin{pmatrix} \alpha_1 \\ \vdots \\ \alpha_p \end{pmatrix}, \quad \beta = \begin{pmatrix} \beta_1 \\ \vdots \\ \beta_p \end{pmatrix} \quad \text{and} \quad \mathbf{e}_t = \begin{pmatrix} e_{1t} \\ \vdots \\ e_{pt} \end{pmatrix} \quad (7)$$

so that (conveniently dropping the index t)

$$E(\mathbf{R}) = \boldsymbol{\alpha} + \beta \mu_I \quad (8)$$

and

$$\text{cov}(\mathbf{e}) = \begin{pmatrix} \sigma_{e1}^2 & 0 & \dots & 0 \\ 0 & \sigma_{e2}^2 & \dots & 0 \\ \vdots & \vdots & \ddots & \vdots \\ 0 & \dots & \dots & \sigma_{ep}^2 \end{pmatrix}. \quad (9)$$

This implies that

$$\text{cov}(\mathbf{R}) = \sigma_I^2 \beta \beta' + \begin{pmatrix} \sigma_{e1}^2 & 0 & \dots & 0 \\ 0 & \sigma_{e2}^2 & \dots & 0 \\ \vdots & \vdots & \ddots & \vdots \\ 0 & \dots & \dots & \sigma_{ep}^2 \end{pmatrix}. \quad (10)$$

Here we have assumed that all e_{it} , $i = 1, \dots, p$, $t = 1, \dots, N$ are independent. We now relax this assumption by assuming that the disturbances e_{it} of the different stocks are dependent (correlated). There is evidence that this is indeed the case on the JSE, but the correlations are not high.

Thus

$$\begin{aligned} E(e_{it}e_{jt}) &= \sigma_{ij}, \quad i \neq j \\ &= \sigma_{ei}^2 \quad i = j \\ &= \sigma_{ii}. \end{aligned}$$

or

$$E(\mathbf{ee}') = \Omega = \begin{pmatrix} \sigma_{e1}^2 & \sigma_{12} & \dots & \sigma_{1p} \\ \sigma_{21} & \sigma_{e2}^2 & \dots & \sigma_{2p} \\ \vdots & \vdots & \ddots & \vdots \\ \sigma_{p1} & \dots & \dots & \sigma_{ep}^2 \end{pmatrix} \quad (11)$$

with

$$\text{cov}(\mathbf{R}) = \sigma_I^2 \beta \beta' + \Omega = \Phi \quad (12)$$

For portfolio $P = W'R$ we have

$$E(P) = \mathbf{W}'(\alpha + \beta\mu_I) = \mu_p \quad (13)$$

and

$$var(P) = W'(\sigma_I^2 \beta \beta' + \Omega)W \quad (14)$$

$$= W' \Phi W = \sigma_p^2 \quad (15)$$

and we maximize

$$Z = \phi \mu_p - \sigma_p^2 \quad (16)$$

subject to

$$\sum_{i=1}^p w_i = 1$$

and any other equality or inequality constraints and bounds. This is again a straightforward QP problem and can be solved in the usual way by using the simultaneous set of equations with Φ replacing Σ . The quantities to be estimated are

$$\mu_I, \sigma_I^2, \alpha, \beta \text{ and } \Omega. \quad (17)$$

To estimate Ω let

$$\hat{\mathbf{E}} = \begin{pmatrix} \hat{e}_{11} & \hat{e}_{12} & \dots & \hat{e}_{1N} \\ \hat{e}_{21} & \hat{e}_{22} & \dots & \hat{e}_{2N} \\ \vdots & \vdots & \vdots & \vdots \\ \hat{e}_{p1} & \dots & \dots & \hat{e}_{pN} \end{pmatrix} \quad (18)$$

then

$$\hat{\Omega} = \frac{1}{N-2} \hat{\mathbf{E}} \hat{\mathbf{E}}'. \quad (19)$$

and

$$\hat{\Phi} = \hat{\sigma}_I^2 \hat{\beta} \hat{\beta}' + \hat{\Omega}. \quad (20)$$

We give below the computed Efficient Frontiers for our model (16) for the nine stocks discussed above. As a comparison we also include the Efficient Frontier for the diagonal

model where

$$E(\mathbf{e}_t \mathbf{e}_t') = \begin{pmatrix} \sigma_{e1}^2 & 0 & \dots & 0 \\ 0 & \sigma_{e2}^2 & \dots & 0 \\ \vdots & \vdots & \ddots & \vdots \\ 0 & \dots & \dots & \sigma_{ep}^2 \end{pmatrix} \quad (21)$$

with estimates the diagonal elements of $\hat{\Phi} = \hat{\sigma}_I^2 \hat{\beta} \hat{\beta}' + \hat{\Omega}$. It is clear that the model with uncorrelated errors, although only marginally, outperforms the model where the correlations between errors of the stocks are assumed to be non-zero. (Hossain (2005))

8.3 Improved Sharpe versus Sharpe - Single Index Case

8.3.1 The Data

For this empirical analysis, the data used was the MIDS batch.

8.3.2 Study Objectives

This chapter illustrates innovations to the current Sharpe Single Index (SSI) model. These formulations are a result of introducing the assumption that the disturbances e_{it} for each of our 9 stocks are correlated, i.e. the Improved Sharpe Single Index (ISSI) Model. The objective of this study will be to recompute efficient frontiers for $GARCH(1, 1)$, $ARMA$, State Space and least squares, employing the Troskie Hossain Innovation, for the single index model. We show that within the ISSI Model, that $GARCH(1, 1)$ is still superior within the South African stock market context. This will be in accordance with the central hypothesis of the thesis. In addition, the study will further demonstrate a comparison between the Sharpe Single Index (SSI) model and the Improved Sharpe Single Index (ISSI) Model.

8.3.3 Methodology

Having computed Sharpe efficient frontiers from before, most of the inputs regarding the different estimation techniques have already been collected. Up to now, we have already estimated the β 's for the index, JSE All share, for all 9 stocks regarding the $GARCH(1, 1)$, $ARMA$ and least square frameworks. The next step is to calculate the new covariance

structure regarding these 3 estimation techniques. Thus

$$\begin{aligned} E(e_{it}e_{jt}) &= \sigma_{ij}, \quad i \neq j \\ &= \sigma_{ei}^2 \quad i = j \\ &= \sigma_{ii}. \end{aligned}$$

or

$$E(\mathbf{ee}') = \mathbf{\Omega} = \begin{pmatrix} \sigma_{e1}^2 & \sigma_{12} & \dots & \sigma_{1p} \\ \sigma_{21} & \sigma_{e2}^2 & \dots & \sigma_{2p} \\ \vdots & \vdots & \ddots & \vdots \\ \sigma_{p1} & \dots & \dots & \sigma_{ep}^2 \end{pmatrix}$$

with

$$\text{cov}(\mathbf{R}) = \sigma_I^2 \beta \beta' + \mathbf{\Omega} = \mathbf{\Phi}$$

Having calculated this new covariance structure for each of the estimation techniques under the current analysis, we then apply the optimisation algorithm on each of the sets of inputs regarding the various modelling techniques. The optimisation program 'MARKO9', is a FORTRAN programme, specifically written for the construction of mean-variance frontiers. It was graciously obtained from Professor C.G. Troskie from the Statistical Sciences Department at the University of Cape Town. Upon obtaining the mean and variance vectors regarding each technique, they are then plotted against each other and the relevant comparisons are then made.

8.3.4 Primary Findings

The above outlined methodology was implemented for the following estimation techniques: *GARCH*(1, 1), *ARMA*, Least Squares (*OLS*), Maximum Likelihood Estimation (*MLE*) and State Space modelling (*SS-models*). Figure 8.1 below illustrates the efficient frontiers

constructed for the 5 aforementioned estimation techniques utilising the Improved Sharpe formulation:

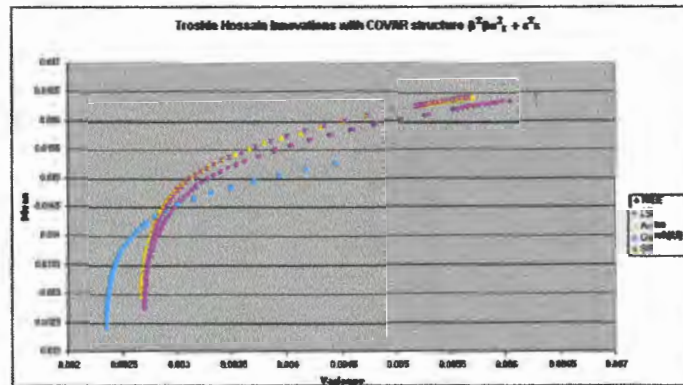


Figure 8.1: Efficient Frontiers for the ISSI Model

For the ISSI model, we find that the $GARCH(1, 1)$ model coincides with the other 4 modelling techniques at a point. To the left of these points of intersection, the $GARCH(1, 1)$ model exhibits superior risk-return results. Whilst the converse holds true to the right of the point of intersection. Moreover, the bulk of the points are on the left hand side of the intersection points, thus overall, $GARCH(1, 1)$ is superior. Comparing the remaining 4 techniques, $ARMA$, OLS , MLE with each other, a consistent pattern is observed similar to the case in the Sharpe Single index model. It is established that OLS is the worst, then follows the State Space and finally the $ARMA$ model outperforming it marginally in the respective order. The MLE and OLS in the Markowitz model yield similar estimates of β hence the efficient frontiers are very close to each other. This confirms to expectations set out at the beginning of the thesis.

Like in the SSI context, the efficient frontiers observed under the ISSI context have produced consistent results with that of their SSI counterparts. We now establish the order of superiority within the ISSI case: it is empirically established that $GARCH(1, 1)$ is the

most superior, then *ARMA*, State Space follows, and finally then *OLS* and *MLE*. The subsequent crucial step is to now compare the SSI frontiers with the ISSI frontiers within the *ARMA*, *GARCH*(1,1), *OLS* and State Space (SS) frameworks. Figures 8.2 , 8.3, 8.4 and Figure 8.5 below compare the SSI and ISSI frontiers in a *GARCH*(1,1), *ARMA*, *OLS* and *SS* framework respectively

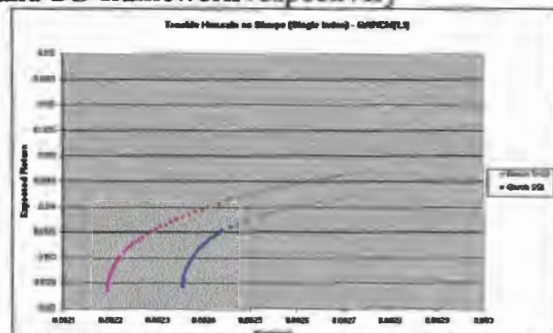


Fig 8.2 Garch SSI versus Garch ISSI

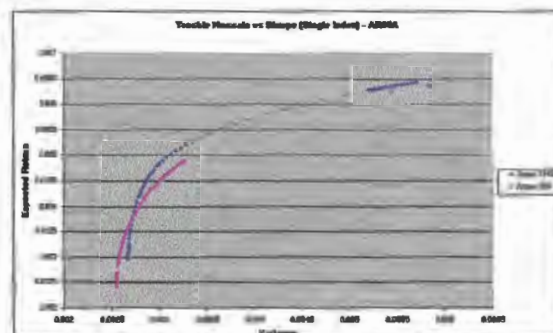


Figure 8.3: Arma SSI versus Arma ISSI

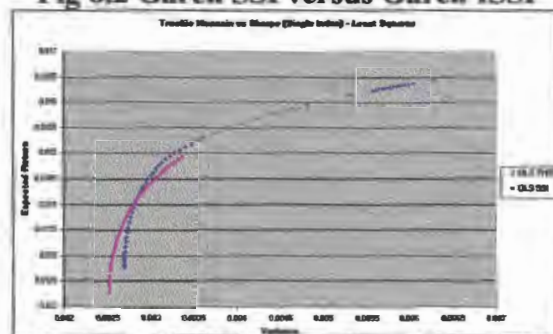


Figure 8.4: OLS SSI versus OLS ISSI

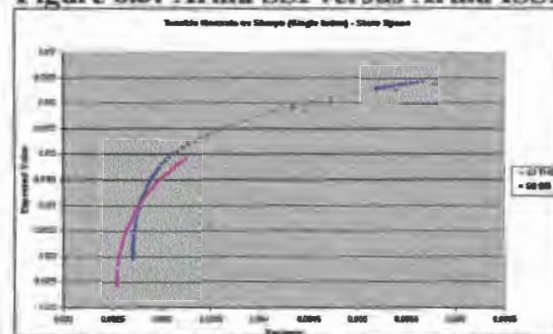


Figure 8.5: SS SSI versus SS ISSI

In figure 8.2, we compare the *GARCH*(1,1) frontier for the SSI and ISSI models.

Clearly the SSI frontier for *GARCH*(1,1) lies to the left of its ISSI counterpart, thus marginally outperforming the ISSI model. From the theory presented in the previous section of this chapter, it is apparent that the ISSI model incorporates more information into its framework than the SSI model through the introduction of the assumption of correlated residuals. Thus intuition tells us that the ISSI model portrays a more accurate and realistic illustration of the market's true risk-return frontier of a particular estimation technique than what Sharpe's model is portraying.

The $GARCH(1,1)$ SSI frontier is to the left of the $GARCH(1,1)$ ISSI frontier, hence it is obvious from the above intuition that Sharpe is underestimating risk at every level of return when compared to Troskie and Hossain. Figures 8.3, 8.4 and 8.5 present a similar scenario for the $ARMA$, SS and OLS framework respectively. Both contain a point of coincidence where the converse holds true to the right of this intersection point. Again the bulk of the risk-return points are contained to the left of the intersection points in 8.3, 8.4 and 8.5. It thus follows again that Sharpe's model exhibits a lower risk level associated with each return platform for the $ARMA$, SS and OLS frameworks.

Hence figures 8.2, 8.3, 8.4 and 8.5 illustrate that Sharpe's model outperforms the ISSI model for all 4 frameworks. Due to the incorporation of more information by the ISSI model, it is also apparent that Sharpe's model under-estimates the portfolio's risk for the 3 given estimation techniques.

8.3.5 Conclusion

Comparing the Sharpe Single Index model against the Improved Sharpe Single Index model for the $GARCH(1,1)$, $ARMA$, SS and OLS estimation techniques, it can be concluded that most of the time the SSI model is under-rating risk compared to the ISSI model. Also, the ISSI model incorporates more information into its model than the SSI model, hence the ISSI model portrays a more accurate and realistic risk-return level for the market than its SSI counterpart. Thus the Sharpe model is more restricted than the Troskie Hossain model. It was also concluded that both SSI and ISSI produced consistent empirical results with respect to the 4 main estimation techniques under study.

The next section compares the Sharpe model with the Model proposed by Troskie and Hossain, The Improved Sharpe, in a multiple index context for the $GARCH(1, 1)$, $ARMA$, OLS and MLE frameworks (Hossain (2005)).

8.4 Improved Sharpe Multiple Index Model.

The multi - index model can be written as

$$\begin{aligned} R_{it} &= \alpha_i + \beta_{i1}I_{1t} + \beta_{i2}I_{2t} + \dots + \beta_{iM}I_{Mt} + e_{it}, \\ i &= 1, \dots, p, \quad t = 1, \dots, N, \end{aligned} \quad (1)$$

with the following assumptions

$$E(e_{it}^2) = \sigma_{ei}^2 \quad (2)$$

$$E(e_{it}e_{jt}) = \sigma_{ij}, \quad i \neq j, \quad t = 1, \dots, N. \quad (3)$$

$$E(e_{it}e_{is}) = 0, \quad t \neq s = 1, \dots, N, \quad (4)$$

$$E(e_{it}I_{jt}) = 0, \quad j = 1, \dots, M, \quad t = 1, \dots, N, \quad (5)$$

$$E(I_{jt}I_{kt}) = c_{jk}, \quad j, k = 1, \dots, M \quad (6)$$

These assumptions are identical to the Single Index Model defined by Troskie and Hossain where in equation (3) we now also assume that the disturbance term e_{it} has a covariance term (correlated) but from assumption (5) remains independent of the Indices I_j , $j = 1, \dots, M$. This last assumption again is the normal assumption in regression. We further

assume that the Indices are dependent with covariances given by c_{jk} (Hossain (2005)). Let

$$E_i = E(R_i) = \alpha_i + \beta_{i1}\mu_1 + \dots + \beta_{iM}\mu_M, \quad (7)$$

$$i = 1, \dots, p.$$

Let

$$\alpha = \begin{pmatrix} \alpha_1 \\ \vdots \\ \alpha_p \end{pmatrix}, \quad \beta = \begin{pmatrix} \beta_{11} & \dots & \beta_{1M} \\ \vdots & \vdots & \vdots \\ \beta_{p1} & \dots & \beta_{pM} \end{pmatrix}, \quad \mathbf{e}_t = \begin{pmatrix} e_{1t} \\ \vdots \\ e_{pt} \end{pmatrix}$$

$$R_t = \begin{pmatrix} R_{1t} \\ \vdots \\ R_{pt} \end{pmatrix}, \quad I_t = \begin{pmatrix} I_{1t} \\ \vdots \\ I_{Mt} \end{pmatrix}, \quad t = 1, \dots, N,$$

then

$$\mathbf{R}_t = \alpha + \beta \mathbf{I}_t + \mathbf{e}_t \quad (8)$$

with

$$E(\mathbf{R}_t) = \alpha + \beta \mu_{\mathbf{I}} \quad (9)$$

and

$$\mu_{\mathbf{I}} = \begin{pmatrix} \mu_1 \\ \vdots \\ \mu_M \end{pmatrix}. \quad (10)$$

The covariance matrix of \mathbf{R}_t is then

$$\begin{aligned} \text{cov}(\mathbf{R}_t) &= E[\mathbf{R}_t - E(\mathbf{R}_t)][\mathbf{R}_t - E(\mathbf{R}_t)]' \\ &= E[\alpha + \beta \mathbf{I}_t + \mathbf{e}_t - \alpha - \beta \mu_{\mathbf{I}}][\alpha + \beta \mathbf{I}_t + \mathbf{e}_t - \alpha - \beta \mu_{\mathbf{I}}]' \\ &= E[\beta(\mathbf{I}_t - \mu_{\mathbf{I}}) + \mathbf{e}_t][\beta(\mathbf{I}_t - \mu_{\mathbf{I}}) + \mathbf{e}_t]' \\ &= \beta E(\mathbf{I}_t - \mu_{\mathbf{I}})(\mathbf{I}_t - \mu_{\mathbf{I}})' \beta' + E(\mathbf{e}_t \mathbf{e}_t') \quad \text{since } E(\mathbf{I}_t \mathbf{e}_t') = 0 \\ &= \beta \mathbf{C} \beta' + \Omega \\ &= \Phi. \end{aligned} \quad (11)$$

For portfolio

$$P = W'R \quad (12)$$

$$= \sum_{i=1}^p w_i R_i \quad (13)$$

we have that

$$E_p = E(P) = W'(\alpha + \beta\mu_I) \quad (14)$$

and

$$\sigma_p^2 = \text{var}(P) = W'\Phi W. \quad (15)$$

The portfolio problem is then

$$\min Z = -\Lambda E_p + \sigma_p^2 \quad (16)$$

$$= -\Lambda W'(\alpha + \beta\mu_I) + W'\Phi W. \quad (17)$$

subject to

$$\sum_{i=1}^p w_i = 1$$

and any other equalities, in-equalities or bounds. This is again a straightforward QP -problem.

If there are no further equalities, in-equalities or bounds, then the problem can be solved using calculus by using the multiple index simultaneous equations with Φ substituted for Σ . Our estimates would be

$$\hat{E}_p = W'\hat{E} \quad (18)$$

where

$$\hat{E} = \begin{pmatrix} \hat{E}_1 \\ \vdots \\ \hat{E}_p \end{pmatrix}$$

can again be estimated in two ways that is

$$\hat{E}_i = \frac{1}{N} \sum_{t=1}^N R_{it} \quad (19)$$

or

$$\hat{E}_i = \hat{\alpha} + \hat{\beta} \hat{\mu}_I \quad (20)$$

where

$$\hat{\alpha} = \begin{pmatrix} \hat{\alpha}_1 \\ \vdots \\ \hat{\alpha}_p \end{pmatrix}, \quad \hat{\beta} = \begin{pmatrix} \hat{\beta}_{11} & \cdots & \hat{\beta}_{1M} \\ \vdots & \vdots & \vdots \\ \hat{\beta}_{p1} & \cdots & \hat{\beta}_{pM} \end{pmatrix} \quad (21)$$

and

$$\hat{\mu}_I = \begin{pmatrix} \hat{\mu}_{I_1} \\ \vdots \\ \hat{\mu}_{I_M} \end{pmatrix} \quad (22)$$

with $\hat{\mu}_I$ some estimate of μ_I (forecast or prediction or Bayesian estimate). If

$$\hat{\mu}_{I_j} = \frac{1}{N} \sum_{t=1}^N I_{jt}.$$

(which is the sample mean of the j^{th} Index) then \hat{E}_i and $\hat{\mu}_I$ are identical. For the estimate of σ_p^2 we have

$$\hat{\sigma}_p^2 = \mathbf{W}' \hat{\Phi} \mathbf{W}$$

with

$$\hat{\Phi} = \hat{\beta} \hat{C} \hat{\beta}' + \hat{\Omega}.$$

The matrix \hat{C} is the estimated covariance matrix of the M Indices and $\hat{\Omega}$ is estimated by

$$\hat{\Omega} = \frac{1}{N - M - 1} \hat{\mathbf{E}} \hat{\mathbf{E}}' \quad (23)$$

and

$$\hat{\mathbf{E}} = \begin{pmatrix} \hat{e}_{11} & \hat{e}_{12} & \cdots & \hat{e}_{1N} \\ \hat{e}_{21} & \hat{e}_{22} & \cdots & \hat{e}_{2N} \\ \vdots & \vdots & \vdots & \vdots \\ \hat{e}_{p1} & \cdots & \cdots & \hat{e}_{pN} \end{pmatrix}. \quad (24)$$

8.5 Improved Sharpe versus Sharpe - Multiple Index Case

8.5.1 The Data

For this empirical analysis, the data used was the MIDS batch.

8.5.2 Study Objectives

Previously it was established that Sharpe under-states risk within the single index context. Within this empirical study we attempt to establish a similar objective within the multiple index framework. Furthermore, we compare the two frameworks, Sharpe and Improved Sharpe, for two portfolio cases. The first case constitutes an environment that exhibits predominantly positive correlated residuals amongst stocks, whilst the contra case exhibits principally negative correlation between the residuals of the respective portfolio. In the first case it is intended to illustrate that Sharpe underestimates risk, whilst for the opposing case for negative correlation we demonstrate the opposite, Sharpe overestimates the risk. The consistency of this claim will be explored for $GARCH(1, 1)$, $ARMA$, OLS and State space multiple index models implementing the Troskie Hossain formulation for the covariance structures of the Indices and the disturbance terms e_{it} .

Again in support of the central hypothesis, it is to be illustrated that the $GARCH(1, 1)$ frontier is still superior within the South African stock market context within the Improved Sharpe Multiple Index (ISMI) Model.

8.5.3 Methodology

Having computed multiple index Sharpe efficient frontiers from before, most of the inputs regarding the different estimation techniques have already been collected. Up to now, we have already estimated the β' s for the 3 indices, JSE, Paladium and Implats, for all 9 stocks regarding the $GARCH(1, 1)$, $ARMA$, State Space and least square frameworks. The next step is to calculate the new covariance structure regarding these 3 estimation techniques. Thus

$$\begin{aligned} E(e_{it}e_{jt}) &= \sigma_{ij}, \quad i \neq j, \quad t = 1, \dots, N. \\ &= \sigma_{ei}^2 \quad i = j \\ &= \sigma_{ii}. \end{aligned}$$

$$E(I_{jt}I_{kt}) = c_{jk}, \quad j, k = 1, \dots, M$$

or

$$E(ee') = \Omega = \begin{pmatrix} \sigma_{e1}^2 & \sigma_{12} & \dots & \sigma_{1p} \\ \sigma_{21} & \sigma_{e2}^2 & \dots & \sigma_{2p} \\ \vdots & \vdots & \ddots & \vdots \\ \sigma_{p1} & \dots & \dots & \sigma_{ep}^2 \end{pmatrix}$$

with

$$cov(\mathbf{R}) = \beta C \beta' + \Omega = \Phi$$

Having calculated this new covariance structure for each of the estimation techniques under the current analysis, we then apply the optimisation program 'MARKO9' on each of the sets of inputs regarding the various modelling techniques. The optimisation program 'MARKO9', is a FORTRAN programme, specifically written for the construction of mean-variance frontiers. It was graciously obtained from Professor C.G. Troskie from the Statis-

tical Sciences Department at the University of Cape Town. Upon obtaining the mean and variance vectors regarding each technique, they are then plotted against each other and the relevant comparisons are then made.

8.5.4 Primary Findings

The above outlined methodology was implemented for the following estimation techniques: *GARCH*(1, 1), *ARMA*, Least Squares (*OLS*) and State Space modelling (SS-models). Figure 8.6 below illustrates the efficient frontiers constructed for the 4 aforementioned estimation techniques utilising the Troskie Hossain (Improved Sharpe) innovation for the multiple index setting, for the positive correlation case:

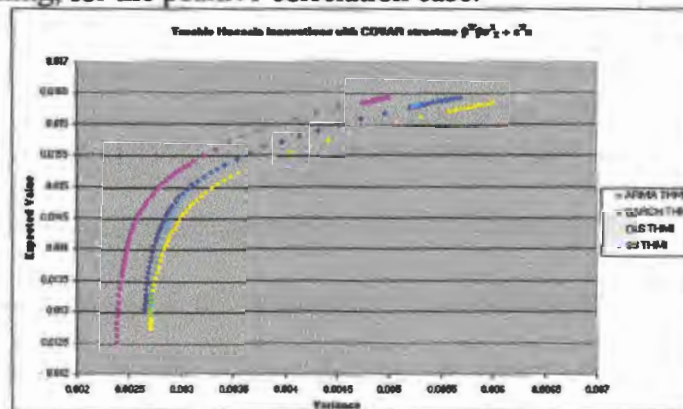


Figure 8.6: Frontiers for the ISMI models

Similar to the single index scenario, the efficient frontiers observed under the ISMI context have produced consistent results with that of their ISMI counterparts. We now establish the order of superiority within the ISMI positive correlation case: it is empirically established that *GARCH*(1, 1) is the most superior, then *ARMA*, State Space follows next, and finally then *OLS*. Next we compare the SMI (Sharpe Multiple Index) frontiers with the ISMI frontiers within the *ARMA*, *GARCH*(1, 1), *OLS* and State Space (SS) frameworks for

the multiple index setting. Figures 8.7, 8.8, 8.9 and Figure 8.10 below compare the SMI and ISMI frontiers in a $GARCH(1, 1)$, $ARMA$, OLS and SS framework respectively.

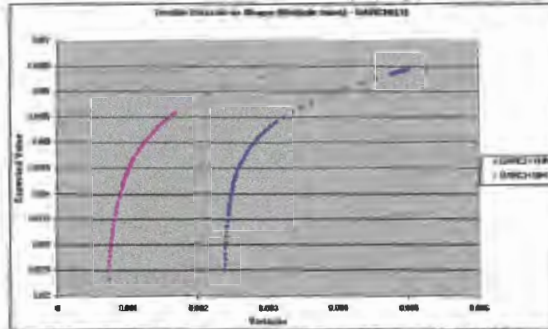


Fig 8.7: Garch SMI vs Garch ISMI

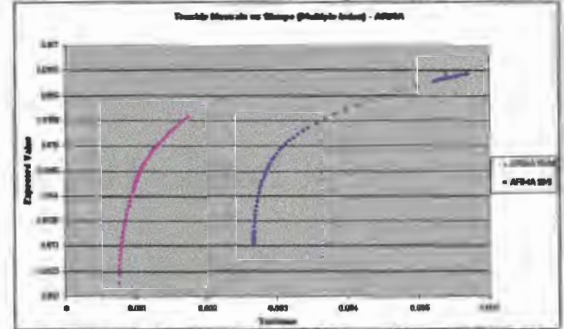


Figure 8.8: Arma SMI vs Arma ISMI

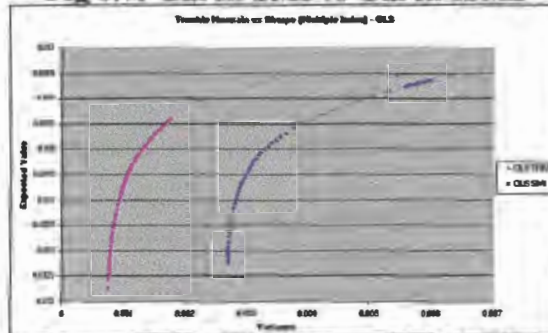


Figure 8.9: OLS SMI vs OLS ISMI

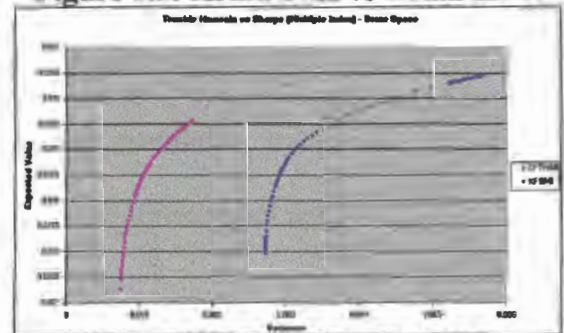


Figure 8.10: SS SMI vs SS ISMI

Like in the single index setting, for the ISMI (Improved Sharpe Multiple Index)

model, we find that the $GARCH(1,1)$ model has the most superior risk-return frontier of the 4 estimation techniques compared, which is again consistent with our central hypothesis. Comparing the remaining 3 techniques, $ARMA$, OLS and State Space with each other in the ISMI environment, a consistent pattern is observed similar to the single index case. It is established that OLS is the worst, then the State Space and $ARMA$ model outperforming it marginally in the respective order.

From the four above comparisons given in figures 8.7 to 8.10, it is clear that in each of the 4 cases, the SMI model has demonstrated a far more superior efficient frontier than the ISMI. Recall that the ISMI is a further improvement on the SMI model. It introduces the assumption that the disturbances e_{it} are correlated since there is evidence that this is indeed

the case on the JSE, but the correlations are not high. Hence providing more information than the the SMI model since. As a result the ISMI model is a more realistic model of the market than the SMI model. Hence in the above 4 graphs it can be established that Sharpe is understating the portfolio's risk even greater than before due to the SMI model ignoring the correlation structure that exists between the residuals (Hossain (2005)).

In the multiple index setting this is even more pronounced due to there being more indices now, hence the SMI model ignores more information than before thus further understating the current risk level. As a result, the ISMI and ISSI models are a more accurate perception of the portfolio's true risk-return profile in the market. This is mainly attributed to the fact that Troskie Hossain Models incorporate more information than the SMI and SSI models and adjust appropriately for the additional risk present from the existing correlation structure.

Next we investigate the case for a portfolio exhibiting a pre-dominantly negatively correlated residual structure. In the above section it was empirically shown using South African market data that for portfolios exhibiting positive correlated residuals, the SMI 'over-estimates' risk when compared with ISMI. We empirically attempt to establish that under the negative correlation case, the SMI framework consistently 'over-estimates' risk when compared with its ISMI counterpart. The methodology adopted to show this empirical result is as follows:

Step 1: Regress the three significant market indices with each stock in the portfolio.

Step 2: Fit relative models of interest for each stock in portfolio i.e. $GARCH(1,1)$ (General Autoregressive Conditional Heteroskedastic), ARMA (Autoregressive Moving Average) models, State Space (SS) models and OLS (Ordinary Least Squares) model.

Step 3: Extract betas, SERs, stock means and respective model residuals for each estimation method

Step 4: Estimate respective covariance structures under SMI and ISMI frameworks

Step 5: Implement Sharpe Algorithm for efficient frontier construction in any programming language

Step 6: Plot standard errors versus returns

Figure 8.11 below illustrates the risk-return frontiers obtained from implementing the above methodology.

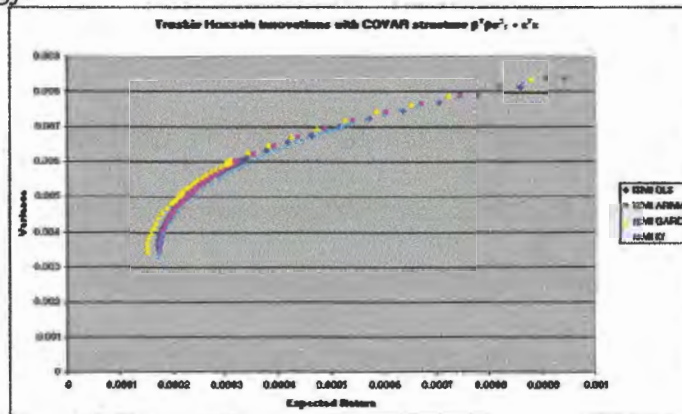


Fig 8.11: Frontiers for the ISMI models - Negative

Immediately it is confirmed from the above illustration in Figure 8.11 of $GARCH(1,1)$'s superiority amongst its rival techniques. Hence the central hypothesis holds. Having confirmed the efficiency and superiority of $GARCH(1,1)$, next we compare the SMI and ISMI models within the negative correlation case.

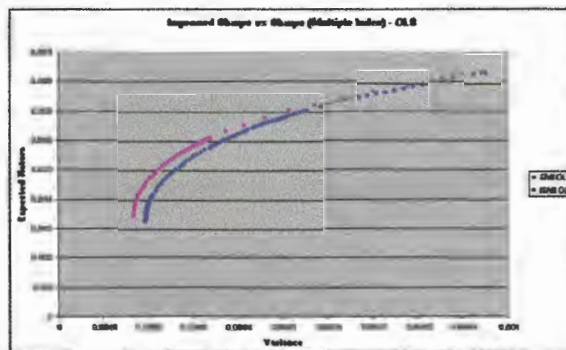


Fig 8.12: OLS SMI vs OLS ISMI

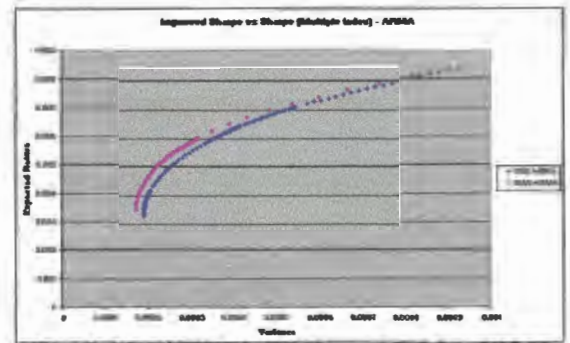


Fig 8.13: ARMA SMI vs ARMA ISMI

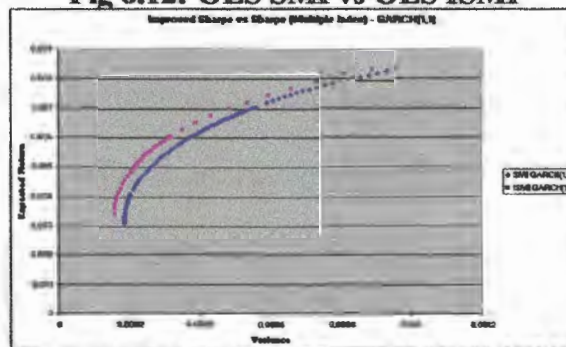


Fig 8.14: GARCH SMI vs GARCH ISMI

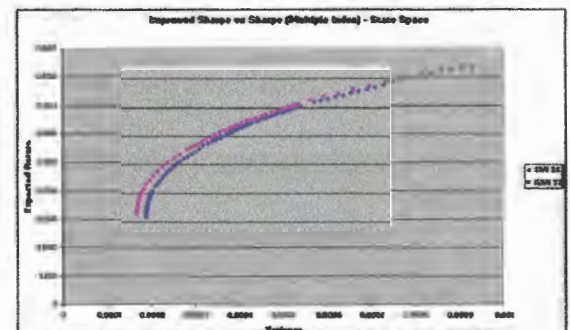


Fig 8.15: SS SMI vs SS ISMI

The consistent relationship that one should observe is that the SMI frontier lies to

the right of the ISMI frontiers. This is the case for all four estimation methods. Hence for any given level of return, the SMI frontiers perceive a higher level of risk associated with the investor's portfolio of stocks than that of the ISMI. Keeping in mind the following important 3 points:

1. that the SMI framework ignores the correlation structure of the portfolio,
2. that SMI is less complete in terms of assumptions and information in contrast to ISMI formulations
3. and finally that negatively correlated asset prices according to logic and financial common sense diversifies away unique risk

Taking note of these 3 points it is easy to ascertain that the ISMI framework is a more accurate paradigm for modelling risk return relationships than its counterpart (Hossain (2006)). This therefore leads us to our conclusion presented in the next section.

8.5.5 Conclusions

Examining the findings from the previous section and considering the theory presented, it is very clear from the mathematical arguments disclosed that the ISMI frontiers are the most accurate perception of the risk and return relationship that can exist for any combination of stocks in an investor's portfolio. As a consequence of the results obtained in the previous section, it can be deduced, from an empirical stand point, for the negative correlation case that the SMI frontier lies consistently to the right of the ISMI frontier. As a consequence of this consistent behaviour portrayed, it is empirically conclusive that the SMI estimation framework is consistently 'over estimating' risk with regards to portfolios of stocks with a pre-dominantly negatively dependent covariance structure (Hossain (2006)). Similarly for the positively correlated case, the opposite was observed, whereby the SMI models were consistently underestimating risk, hence its frontiers were lying consistently to the left of their ISMI counterparts (Hossain (2005)).

The rational behind this is that the SMI covariance structure ignores the presence of covariance dependencies (i.e. off diagonal elements equal 0), thus the SMI efficient frontier will always lie to the right of the ISMI frontier in the case of portfolios exhibiting pre-dominantly negatively correlated residuals. In the case of portfolios demonstrating positive correlation, one should find that the SMI frontier will lie to the left of the ISMI

frontier. In addition it is concluded that the original formulated ISMI multiple index model is in fact a more accurate risk evaluation method in the presence of correlated residuals. This is to be verified by a principal component analysis study that is to be conducted the next chapter.

Chapter 9

Principal Components Analysis

9.1 Introduction

Finally we come to Principal Components Analysis (PCA). What is it? It is a way of identifying patterns in data, and expressing the data in such a way as to highlight their similarities and differences. Since patterns in data can be hard to find in data of high dimension, where the luxury of graphical representation is not available, PCA is a powerful tool for analysing data. The other main advantage of PCA is that once you have found these patterns in the data you compress the data by reducing the number of dimensions, without much loss of information.

Principal components are linear combinations of random or statistical variables which have special properties in terms of variances. For example, the first principal component is the normalised linear combination (that is, the sum of squares of the coefficients being one) with maximum variance. In effect, transforming the original vector variable to the vector of principal components amounts to a rotation of coordinate axes to a new coordinate system that has inherent statistical properties. This choosing of a coordinate system is to be contrasted with the many problems treated previously where the coordinate system is irrelevant.

The principal components turn out to be the characteristic vectors of the covariance matrix. Thus the study of principal components can be considered as putting into statistical

terms the usual developments of characteristic roots and vectors (for positive semi-definite matrices). From the point of view of statistical theory, the set of principal components yields a convenient set of coordinates, and the accompanying variances of the components characterise their statistical properties. In statistical practice, the method of principal components is used to find the linear combinations with large variance. In many exploratory studies the number of variables under consideration is too large to handle. Since it is the deviations in these studies which are of interest, a way of reducing the number of variables to be treated is to discard the linear combinations which have small variances and study only those with large variances.

The main use of PCA is to reduce the dimensionality of a data set while retaining as much information as is possible. It computes a compact and optimal description of the data set. The first principal component is the combination of variables that explains the greatest amount of variation. The second principal component defines the next largest amount of variation and is independent to the first principal component. There can be as many possible principal components as there are variables. It can be viewed as a rotation of the existing axes to new positions in the space defined by the original variables. In this new rotation, there will be no correlation between the new variables defined by the rotation. The first new variable contains the maximum amount of variation, the second new variable contains the maximum amount of variation unexplained by the first and orthogonal to the first, etc...

There are several algorithms for calculating the Principal Components. Given the same starting data they will produce the same results with the one exception (are you sur-

prised?). This exception is that, if at some point, there are two or more possible rotations that contain the same "maximum" variation, then which one is used is indeterminate. In two dimensions the data cloud would look like a circle, instead of an ellipse. In a circle, any rotation would be equivalent. In an elliptical data cloud, the first component would be parallel to the major axis of the ellipse.

It can be viewed as finding a projection of the observations onto orthogonal axes contained in the space defined by the original variables. The criteria being that the first axis "contains" the maximum amount of variation, or "accounts" for the maximum amount of variation. The second axis contains the maximum amount of variation orthogonal to the first. The third axis contains the maximum amount of variation orthogonal to the first and second axis and so on until one has the last new axis which is the last amount of variation left. As you can see these are really two slightly different ways of saying the same thing!

This chapter is organised as follows, in the next section we present the major theorems and vital mathematical theory and properties required for carrying out principal components analysis. Following that is some theoretical background regarding the Singular Value Decomposition (SVD) of a matrix, where it is theoretically explored as to the construction of the actual orthogonal components of a data matrix. Next we explore the algorithm of PCA and summarise the objectives of the analysis. Finally we look at PC in a regression framework which is followed by an empirical study that looks at the unique application of PC using regression in a risk-return setting that is used to back up empirical claims in previous studies suggesting that the Sharpe formulation is in fact under-estimating risk in the presence of a positively correlated residual structure. The empirical study also

backs up previous empirical results regarding the accuracy of the Improved Sharpe formulation as well.(Troskie (2000))

9.2 Some Mathematical Background

Much of the theory on applications that follow are known to most readers (See Anderson (1958)). What is generally no known is that Principal Components applied to stock market data allows one to construct proxies to stock market indices. These new indices captures the variability of stock return movements more efficiently than any other methods known. For example the first principal component constructed will be the index with the largest variance and that means the most information. Furthermore the other principal components are orthogonal (uncorrelated)

Theorem 1.

Let the p -component random vector X have $E(X) = 0$ and $E(XX') = \Sigma$. Then there exists an orthogonal linear transformation

$$U = \beta'X \quad (1)$$

such that the covariance matrix of U is

$$\begin{aligned} E(UU') &= \Lambda \\ &= \begin{pmatrix} \lambda_1 & 0 & \dots & 0 \\ 0 & \lambda_2 & \dots & 0 \\ \vdots & \vdots & \ddots & \vdots \\ 0 & 0 & \dots & \lambda_p \end{pmatrix} \end{aligned} \quad (2)$$

where $\lambda_1 \geq \lambda_2 \geq \dots \geq \lambda_p > 0$ are the roots of

$$|\Sigma - \lambda I| = 0$$

and

$$\beta' \beta = \mathbf{I}, \quad \beta' \Sigma \beta = \Lambda. \quad (3)$$

The r -th column of β , $\beta^{(r)}$, satisfies

$$(\Sigma - \lambda_r \mathbf{I}) \beta^{(r)} = 0.$$

The r -th component, $U_r = \beta^{(r)'} \mathbf{X}$, called the r -th principal component, has maximum variance of all normalized linear combinations uncorrelated with U_1, U_2, \dots, U_{r-1} .

Proof.

To prove the Theorem it is sufficient to derive the first two principal components and conclude with something like *the previous two steps can be repeated p times* which proves the theorem. Note: Equation (3) is often used in Statistics to either, derive distributions (like those of Quadratic Forms), or evaluate multivariate integrals (like the multivariate normal). The first 3 principal components are important and often used for practical purposes. This is because of the variance they explain as a proportion of the Total Variance.

$$\begin{aligned} \text{Total Variance} &= \sum_{i=1}^p \sigma_{ii} \\ &= \text{tr} \Sigma \\ &= \text{tr} \beta \Lambda \beta' \\ &= \text{tr} \Lambda \\ &= \sum_{i=1}^p \lambda_i. \end{aligned}$$

The percentage of the total variation explained by the first three components is given by

$$\frac{\lambda_1 + \lambda_2 + \lambda_3}{\sum_{i=1}^p \lambda_i} 100\%. \quad (4)$$

Theorem 2.

Let $\mathbf{x}_1, \dots, \mathbf{x}_N$ be $N(> p)$ observations from $N(\boldsymbol{\mu}, \boldsymbol{\Sigma})$, where $\boldsymbol{\Sigma}$ is a matrix with p different characteristic roots (eigenvalues). Then a set of maximum likelihood estimates of $\lambda_1 > \lambda_2 > \dots > \lambda_p > 0$ and of $\beta^{(1)}, \dots, \beta^{(p)}$ defined in Theorem 1 are the roots

$$k_1 > k_2 > \dots > k_p$$

of

$$|\hat{\boldsymbol{\Sigma}} - k\mathbf{I}| = 0$$

and the set of corresponding vectors $b^{(1)}, \dots, b^{(p)}$ satisfying

$$(\hat{\boldsymbol{\Sigma}} - k_r \mathbf{I})b^{(r)} = 0$$

and

$$b^{(r)'} b^{(r)} = 1$$

where $\hat{\boldsymbol{\Sigma}}$ is the maximum likelihood estimate of $\boldsymbol{\Sigma}$.

Proof.

Since the roots of $|\boldsymbol{\Sigma} - \lambda\mathbf{I}| = 0$ are all different, each vector $\beta^{(i)}$ is uniquely defined except that $\beta^{(i)}$ can be replaced by $(-\beta^{(i)})$. If we require that the first non-zero component of $\beta^{(i)}$ be positive, then $\beta^{(i)}$ is uniquely defined, and $\boldsymbol{\mu}, \boldsymbol{\Lambda}, \boldsymbol{\beta}$ is a single-valued function of $\boldsymbol{\mu}, \boldsymbol{\Sigma}$. Hence by the uniqueness properties of maximum likelihood estimates the set of maximum likelihood estimates of $\boldsymbol{\mu}, \boldsymbol{\Lambda}, \boldsymbol{\beta}$ are the same functions of $\hat{\boldsymbol{\mu}}, \hat{\boldsymbol{\Sigma}}$. Before we proceed with the computation we list some useful results. The j -th principal component

is

$$\begin{aligned}
 U_j &= \beta^{(j)'} \mathbf{X} \\
 \text{var}(U_j) &= \lambda_j \\
 \text{cov}(U_j U_k) &= 0, \quad j \neq k. \\
 \text{cov}(\mathbf{X} U_j) &= \text{cov}(\mathbf{X} \mathbf{X}' \beta^{(j)}) \\
 &= \Sigma \beta^{(j)} \\
 &= \lambda_j \beta^{(j)} \quad \text{since} \\
 &\quad (\Sigma - \lambda_j) \beta^{(j)} = 0
 \end{aligned}$$

Thus the covariance between X_i (the i -th variable) and U_j is

$$\begin{aligned}
 \text{cov}(X_i U_j) &= \lambda_j \beta_{ij} \quad \text{where} \\
 \beta^{(j)} &= \begin{pmatrix} \beta_{1j} \\ \vdots \\ \beta_{pj} \end{pmatrix}.
 \end{aligned}$$

The correlation between X_i and U_j is

$$\begin{aligned}
 \text{corr}(X_i, U_j) &= \frac{\lambda_j \beta_{ij}}{\sqrt{\sigma_{ii} \lambda_j}} \\
 &= \frac{\beta_{ij}}{\sigma_i} \cdot \sqrt{\lambda_j}
 \end{aligned}$$

so that the weight (loading) β_{ij} makes an important contribution to the covariance and correlation between X_i and U_j . The same holds for the sample (m.l.e.) estimates

$$\begin{aligned}
 \text{cov}(X_i U_j) &= \lambda_j b_{ij} \quad \text{where} \\
 b^{(j)} &= \begin{pmatrix} b_{1j} \\ \vdots \\ b_{pj} \end{pmatrix}
 \end{aligned}$$

and

$$\begin{aligned} \text{corr}(X_i, U_j) &= \frac{k_j b_{ij}}{\sqrt{\hat{\sigma}_{ii} k_j}} \\ &= \frac{b_{ij}}{s_i} \cdot \sqrt{k_j}. \end{aligned}$$

These quantities play an important part in factor analysis.

Sample Computation.

Since $\mathbf{x}_1, \dots, \mathbf{x}_N$ are $N(> p)$ (sample) observations from $N(\boldsymbol{\mu}, \boldsymbol{\Sigma})$ the maximum likelihood estimate of $\boldsymbol{\Sigma}$ is

$$\hat{\boldsymbol{\Sigma}} = \frac{1}{N} \mathbf{A} \quad (5)$$

$$\begin{aligned} &= \frac{1}{N} \sum_{\alpha=1}^N (\mathbf{x}_\alpha - \bar{\mathbf{x}})(\mathbf{x}_\alpha - \bar{\mathbf{x}})' \\ &= \frac{1}{N} \left[\sum_{\alpha=1}^N (x_{i\alpha} - \bar{x}_i)(x_{j\alpha} - \bar{x}_j) \right], \end{aligned} \quad (6)$$

$$i, j = 1, \dots, p \quad (7)$$

$$= \frac{1}{N} [a_{ij}],$$

$$i, j = 1, \dots, p. \quad (8)$$

$$= [\hat{\sigma}_{ij}],$$

$$i, j = 1, \dots, p. \quad (9)$$

The unbiased estimate

$$\mathbf{S} = \frac{1}{N-1} \mathbf{A} \quad (10)$$

may be preferred. From Theorem 2 the sample estimates are

$$\mathbf{B}'\hat{\Sigma}\mathbf{B} = \mathbf{K} \text{ or} \quad (11)$$

$$\hat{\Sigma} = \mathbf{B}\mathbf{K}\mathbf{B}' \text{ where} \quad (12)$$

$$\mathbf{B}'\mathbf{B} = \mathbf{I}$$

that is, \mathbf{B} is orthogonal, and

$$\mathbf{K} = \begin{pmatrix} k_1 & 0 & \dots & 0 \\ 0 & k_2 & \dots & 0 \\ \vdots & \vdots & \ddots & \vdots \\ 0 & 0 & \dots & k_p \end{pmatrix} \quad (13)$$

and k_i are the sample (m.l.e.) eigenvalues or roots. We want to compute \mathbf{B} and \mathbf{K} . The best way to compute them is by using the SVD of a matrix.

Properties of the Covariance Matrix

- The covariance matrix is real and symmetric,
- Hence it is normal, and positive.
- Hence has non-negative eigenvalues
- Hence there exists a unitary matrix \mathbf{B} and a real diagonal matrix \mathbf{K} such that

$$\Sigma = \mathbf{B}^{-1}\mathbf{K}\mathbf{B}$$

- In fact, \mathbf{B} can be chosen to be an orthogonal matrix

Geometric interpretation of \mathbf{B}

\mathbf{B} transforms d dimensional space such that, after transformation

- the off-diagonal entries of the covariance matrix are zero

- the diagonal elements of \mathbf{K} , the eigenvalues of Σ , are the variances of the transformed data along the principal axes of the transformed space
- These eigenvectors are called the principal components of Σ

9.3 The Singular Value Decomposition - SVD - of a Matrix.

Consider any $(n \times p)$ data matrix

$$\mathbf{X}_{(n \times p)} = \begin{pmatrix} X_{11} & \dots & X_{1p} \\ X_{21} & \dots & X_{2p} \\ \vdots & \vdots & \vdots \\ X_{n1} & \dots & X_{np} \end{pmatrix}. \quad (14)$$

Here the rows are the observations and the columns the variables. They can be labeled with names (as usual in the first row for cases and the first column for variables).

The SVD is

$$\mathbf{X} = \mathbf{U}\mathbf{D}\mathbf{V}' \text{ where}$$

$$\mathbf{U}_{(n \times p)} \text{ is orthogonal}$$

$$\mathbf{U}'\mathbf{U} = \mathbf{I}$$

$$\mathbf{D}_{(p \times p)} = \text{diag}(d_1, \dots, d_p),$$

$$d_1 \geq d_2 \geq \dots d_p > 0,$$

$$\mathbf{V}_{(p \times p)} \text{ is orthogonal}$$

$$\mathbf{V}'\mathbf{V} = \mathbf{I}.$$

(Note: The \mathbf{U} used above is a matrix and not the same as vector U used in equation (11).)

Then

$$\mathbf{X}'\mathbf{X} = \mathbf{V}\mathbf{D}\mathbf{U}'\mathbf{U}\mathbf{D}\mathbf{V}' \quad (15)$$

$$= \mathbf{V}\mathbf{D}^2\mathbf{V}'. \quad (16)$$

Notice the similarity between equation (13) and (9). We use this to compute the Principal Components of a sample.

We start off with our data matrix given by

$$\mathbf{X}_{(n \times p)} = \begin{pmatrix} X_{11} & \dots & X_{1p} \\ X_{21} & \dots & X_{2p} \\ \vdots & \vdots & \vdots \\ X_{n1} & \dots & X_{np} \end{pmatrix}$$

.We compute the column means

$$\bar{X}_j = \frac{1}{n} \sum_{i=1}^n X_{ij}, \quad j = 1, \dots, p. \quad (17)$$

We subtract from each element of \mathbf{X} its respective mean i.e.

$$\tilde{\mathbf{X}}_{(n \times p)} = \begin{pmatrix} X_{11} - \bar{X}_1 & \dots & X_{1p} - \bar{X}_p \\ X_{21} - \bar{X}_1 & \dots & X_{2p} - \bar{X}_p \\ \vdots & \vdots & \vdots \\ X_{n1} - \bar{X}_1 & \dots & X_{np} - \bar{X}_p \end{pmatrix} \quad (18)$$

and we call this new matrix $\tilde{\mathbf{X}}$ the *centered matrix*. We divide each element of this matrix by either \sqrt{n} or by $\sqrt{n-1}$ depending on the personal choice of the experimenter.

Thus we have

$$\tilde{\mathbf{X}}_n = \frac{1}{\sqrt{n}} \tilde{\mathbf{X}} \quad (19)$$

and

$$\begin{aligned}
 \hat{\Sigma} &= \left[\frac{1}{\sqrt{n}} \tilde{\mathbf{X}} \right]' \left[\frac{1}{\sqrt{n}} \tilde{\mathbf{X}} \right] \\
 &= \frac{1}{n} \tilde{\mathbf{X}}' \tilde{\mathbf{X}} \\
 &= \frac{1}{n} \mathbf{A} \\
 &= \hat{\Sigma}.
 \end{aligned} \tag{20}$$

the maximum likelihood estimate of the covariance matrix Σ (under multivariate normal theory). The matrix \mathbf{A} is the Wishart matrix (that is the sums of squares and cross products of the centered matrix $\tilde{\mathbf{X}}$

$$\mathbf{A} = \tilde{\mathbf{X}}' \tilde{\mathbf{X}}$$

The SVD is then performed on the adjusted centered matrix

$$\tilde{\mathbf{X}}_n = \frac{1}{\sqrt{n}} \tilde{\mathbf{X}} \tag{21}$$

so that

$$\tilde{\mathbf{X}}_n = \mathbf{U} \mathbf{D} \mathbf{V}' \tag{22}$$

with the result that

$$\begin{aligned}
 \hat{\Sigma} &= \tilde{\mathbf{X}}_n' \tilde{\mathbf{X}}_n \\
 &= \mathbf{V} \mathbf{D}^2 \mathbf{V}' \\
 &= \mathbf{V} \mathbf{K} \mathbf{V}' \\
 &= \mathbf{B} \mathbf{K} \mathbf{B}'
 \end{aligned}$$

with $\mathbf{V} = \mathbf{B}$. It is common practice to use \mathbf{V} instead of \mathbf{B} and we will do the same in all our remaining calculations. If the unbiased estimate $\mathbf{S} = \frac{1}{n-1} \mathbf{A}$ is of interest (as is often

the case) then replace in the calculations $\tilde{\mathbf{X}}_n$ with $\tilde{\mathbf{X}}_{n-1}$. The interpretations will be the same. Thus

$$\tilde{\mathbf{X}}_{n-1} = \mathbf{U}\mathbf{D}\mathbf{V}'$$

and

$$\begin{aligned} \mathbf{S} &= \tilde{\mathbf{X}}_{n-1}' \tilde{\mathbf{X}}_{n-1} \\ &= \mathbf{V}\mathbf{D}^2\mathbf{V}' \\ &= \mathbf{V}\mathbf{K}\mathbf{V}' \\ &= \mathbf{B}\mathbf{K}\mathbf{B}' \end{aligned}$$

The vectors of

$$\mathbf{V} = (\mathbf{V}_1, \mathbf{V}_2, \mathbf{V}_3, \dots, \mathbf{V}_p) \quad (23)$$

are the sample (maximum likelihood estimates) of the characteristic vectors (principal components) and the diagonal elements of

$$\mathbf{K} = \begin{pmatrix} k_1 & 0 & \dots & 0 \\ 0 & k_2 & \dots & 0 \\ \vdots & \vdots & \ddots & \vdots \\ 0 & 0 & \dots & k_p \end{pmatrix} \quad (24)$$

are the sample (m.l.e.) of the characteristic roots (eigenvalues). The sample *score matrix* is then computed as

$$\mathbf{PC}_{score} = \mathbf{PC}_{(n \times p)} = \tilde{\mathbf{X}}_n \mathbf{V} = \tilde{\mathbf{X}}_n (\mathbf{V}_1, \mathbf{V}_2, \mathbf{V}_3, \dots, \mathbf{V}_p). \quad (25)$$

Note that some researchers (and packages like GENSTAT) prefer to use sums of squares and cross product matrix compute $\tilde{\mathbf{X}}$

$$\mathbf{PC}_{(n \times p)} = \tilde{\mathbf{X}} \mathbf{V} = \tilde{\mathbf{X}} (\mathbf{V}_1, \mathbf{V}_2, \mathbf{V}_3, \dots, \mathbf{V}_p). \quad (26)$$

in their calculations where $A = \tilde{X}'\tilde{X}$ is then the Wishart matrix. The interpretation remains the same since there is only a scale difference of $\sqrt{1/n}$ between the two. **NOTE: Most recent research indicates a preference for \tilde{X}_{n-1}** , so that $S = \tilde{X}_{n-1}'\tilde{X}_{n-1}$ is the preferred matrix. We will adopt this procedure later. In equation (22) we often only compute the first three score vectors

$$(\text{PC}_1, \text{PC}_2, \text{PC}_3) = \tilde{X}_n(\mathbf{V}_1, \mathbf{V}_2, \mathbf{V}_3) \quad (27)$$

and these are then plotted in 2-dimensional or 3-dimensional space as projections (representations) of \tilde{X}_n in lower dimensions. The case labels or numbers are used for identification. The case labels are often preferred. This plot is particularly helpful to identify outliers (influential observations). For example we can plot

$$\text{PC}_1 \text{ vs } \text{PC}_2 \quad (28)$$

$$\text{PC}_1 \text{ vs } \text{PC}_3$$

$$\text{PC}_2 \text{ vs } \text{PC}_3$$

$$\text{PC}_1 \text{ vs } \text{PC}_2 \text{ vs } \text{PC}_3. \quad (29)$$

The clarity of a plot can then be expressed as a percentage., Thus

$$\frac{k_1 + k_2}{\sum_{i=1}^p k_i} 100\%$$

is the percentage variation (clarity) explained by the first two component scores. This construction means that we can now see in two dimensions something of what really happens in our original p -dimensions. Hopefully the clarity in 3-dimensions

$$\frac{k_1 + k_2 + k_3}{\sum_{i=1}^p k_i} 100\%$$

will explain a large percentage of our p -dimensional variation. The SVD also allows us to give a geometric view of principal components. Suppose our data matrix (in what form does not matter)

$$\mathbf{X}_{(n \times p)} = \begin{pmatrix} X_{11} & \dots & X_{1p} \\ X_{21} & \dots & X_{2p} \\ \vdots & \vdots & \vdots \\ X_{n1} & \dots & X_{np} \end{pmatrix}$$

can be thought of as giving the coordinates of n points - one for each row - in a multi-dimensional space of p dimensions. It is hard to think of p dimensions - we can only see in three dimensions - and even this is hard enough. One would like to simplify the p dimensions by asking whether the points really lie in p dimensions and, if indeed so, is there a space of fewer dimensions close enough for us to see what really happens. Suppose this space close enough is in two, or three, dimensions, then we can in fact plot the n points in these two or three dimensions to see what really happens.

Mathematically speaking we want to find the best space of low dimensions that is closest to the original space of p dimensions. Here we mean we want to minimize the distance from our p dimensional space to this new space of lower dimensions - which is equivalent of minimizing the squared residuals from the p space, to this new space of lower dimensions. The SVD gives a solution to this problem. Let

$$\begin{aligned} \mathbf{X} &= \mathbf{U}\mathbf{D}\mathbf{V}' \\ &= \mathbf{U}_{[r]}\mathbf{D}_{[r]}\mathbf{V}'_{[r]} + \mathbf{U}_{[p-r]}\mathbf{D}_{[p-r]}\mathbf{V}'_{[p-r]} \\ &= \mathbf{X}_{[r]} + \mathbf{E} \end{aligned} \tag{30}$$

where E , is the residual then $X_{[r]}$ minimizes

$$tr[(X - A)(X - A)'] = \sum_{i=1}^n \sum_{j=1}^p (x_{ij} - a_{ij})^2 \quad (31)$$

among all other $(n \times p)$ matrices A of rank r . Thus the singular value decomposition matrix $X_{[r]}$ can be used as a *matrix approximation* to X . But this is precisely what principal components analysis is doing since

$$\text{Total variance} = \text{Explained variance} + \text{Residual}$$

PCA maximizes the explained variance which is equivalent to minimizing the residual. For the SVD matrix approximation $X_{[3]}$, the clarity is again measured (in three dimensions) by

$$\frac{d_1^2 + d_2^2 + d_3^2}{\sum_{i=1}^p d_i^2} 100\% = \frac{k_1 + k_2 + k_3}{\sum_{i=1}^p k_i} 100\%.$$

Note that some researchers often also plot all the pairs (PC_i, PC_j) and all the triplets (PC_i, PC_j, PC_k) . Many of them see very interesting (amusing) patterns and all sorts of conclusions are made. The small sets of scores

$$(PC_{p-1}, PC_p) \quad (32)$$

are often used to identify collinearities among the explanatory (x_1, x_2, \dots, x_p) variables in regression analysis. This issue is dealt with in more advanced courses. It is also useful to plot the characteristic vectors (often called loadings or factor loadings)

$$(V_1, V_2, V_3, \dots, V_p)$$

in one, two or three dimensions. They often reveal interesting properties amongst the variables - the columns of the data matrix X (or \tilde{X}). (Troskie (2000))

where E_r is the residual then $X_{[r]}$ minimizes

$$\text{tr}[(X - A)(X - A)'] = \sum_{i=1}^n \sum_{j=1}^p (x_{ij} - a_{ij})^2 \quad (31)$$

among all other $(n \times p)$ matrices A of rank r . Thus the singular value decomposition matrix $X_{[r]}$ can be used as a *matrix approximation* to X . But this is precisely what principal components analysis is doing since

$$\text{Total variance} = \text{Explained variance} + \text{Residual}$$

PCA maximizes the explained variance which is equivalent to minimizing the residual. For the SVD matrix approximation $X_{[3]}$, the clarity is again measured (in three dimensions) by

$$\frac{d_1^2 + d_2^2 + d_3^2}{\sum_{i=1}^p d_i^2} 100\% = \frac{k_1 + k_2 + k_3}{\sum_{i=1}^p k_i} 100\%.$$

Note that some researchers often also plot all the pairs (PC_i, PC_j) and all the triplets (PC_i, PC_j, PC_k) . Many of them see very interesting (amusing) patterns and all sorts of conclusions are made. The small sets of scores

$$(PC_{p-1}, PC_p) \quad (32)$$

are often used to identify collinearities among the explanatory (x_1, x_2, \dots, x_p) variables in regression analysis. This issue is dealt with in more advanced courses. It is also useful to plot the characteristic vectors (often called loadings or factor loadings)

$$(V_1, V_2, V_3, \dots, V_p)$$

in one, two or three dimensions. They often reveal interesting properties amongst the variables - the columns of the data matrix X (or \tilde{X}). (Troskie (2000))

9.4 Summary of Objectives and Benefits of PCA

Principal component analysis (PCA) is concerned with explaining the variance covariance structure of the data through a few linear combinations of the original variables. Its general objectives are:

- (a) data reduction
- (b) interpretation.

(a) Data reduction: Although the original data set contains p variables, often much of the variability can be accounted for by a smaller number m of principal components. When there is (almost) as much information in the m components as there is in the original p variables, the original data set consisting of n observations on p variables can be reduced to one consisting of n observations on m principal components.

(b) Interpretation: A PCA can show relationships that were not previously suspected, and it allows interpretations that would not ordinarily result.

The benefits of PCA can be summarised as follows:

Huge computational benefits if one is trying to perform a statistical analysis of the data. E.g.: evaluation of the multinomial Gaussian distribution (i.e. the Gaussian (or normal) distribution extended to multidimensional space) involves inversion of the covariance matrix.

The axis corresponding to the largest eigenvalue is the direction of maximum variation of the data

Hence it is normal in PCA to arrange the basis of

eigenvectors so that the corresponding eigenvalues are in decreasing order.

Magnitude of the higher eigenvalues gives clues to the true underlying dimensionality of Y .

If one is trying to separate the data into different classes, it is natural to start by looking along the axis corresponding to the largest eigenvalue

9.5 Algorithm of PCA

Principal component analysis is straightforward as long as you have a way of computing the eigenvalues and eigenvectors of a covariance matrix. The procedure roughly followed in Eviews 3.1 is:

Step 1. Normalize data to mean zero and variance 1 and divide the data by the square root of the number of observations minus 1 in case a correlation matrix analysis is requested. Center data to have mean zero computing each variable minus its mean in case a covariance matrix is requested and divide the data by the square root of the number of observations minus 1 in case a covariance matrix analysis is requested.

Step 2. Compute the singular value decomposition of the data. This results in the left and right singular vectors and the singular values of the data. Eviews 3.1 uses a function @SVD to carry out this computation.

Step 3. Computation of component scores. This is obtained by multiplying the matrix of left singular vectors by the singular values.

Step 4. Computation of eigenvalues. This is computed squaring the singular values obtained in step 2.

Step 5. Computation of coefficients (eigenvectors). This is the matrix of right singular vectors obtained in step 2.

9.6 Principal Components and Regression.

Principal components can be used in regression where

$$Y = X\beta + e$$

and the principal components are computed from either the X matrix using the SVD or from the

$$[Y, X]$$

matrix using the SVD. It has been demonstrated that PCA in regression can give very satisfactory results, especially, in the presence of collinearity in the explanatory (X) variables.

Comment :

In the next section we will be dealing with an empirical study regarding principal component analysis. As mentioned before a principal component index will be most informative since it captures the maximum variance. The second component also captures the second largest variance but is, of course, uncorrelated (orthogonal) with the first index. The PC1 Index is the solution to

$$\max \text{var}(l'X) = l'\Sigma l \text{ subject to } l'l = 1$$

and the PC2 Index the solution to

$$\max \text{var}(l'X) = l'\Sigma l \text{ subject to } l'l = 1$$

but must be uncorrelated with PC1. Thus if the solutions are

$$PC1 = l'_1 X, \quad l'_1 l_1 = 1 \quad (36)$$

$$PC2 = l'_2 X, \quad l'_2 l_2 = 1, \quad \text{then}$$

$$l'_1 l_2 = 0.$$

If Σ is estimated by $\hat{\Sigma}$ then we compute

$$(\hat{\Sigma} - k_i \mathbf{I}) l_i = 0, \quad i = 1, 2.$$

If \mathbf{x}_α is the α^{th} data point then the PC score is

$$PC1_\alpha = l'_1 \mathbf{x}_\alpha$$

$$PC2_\alpha = l'_2 \mathbf{x}_\alpha, \quad \alpha = 1, \dots, N.$$

In the empirical study to follow, we plot $PC1$, $PC2$, and Jse-Over against Date given by $\alpha = 1, \dots, N$. We will see in the next section that we may add constraints. Mutual Fund managers often require that for one or both Indices

$$l_i \geq 0, \quad i = 1, \dots, p$$

that is, no short-sales are permitted. Such an Index can easily be constructed (see also Affleck-Graves, Money and Troskie, 1979.).

9.7 PCA Study: SMI vs ISMI - Multiple Index Case

9.7.1 The Data

For this empirical analysis, the data set used was the MIDS batch.

9.7.2 Study Objectives

The purpose of this study is to investigate the behavior of efficient frontiers under two developed multiple index portfolio frameworks, namely the Sharpe Multiple Index (abbreviated as SMI) and the Improved Sharpe Multiple Index (abbreviated as ISMI) portfolio models proposed by Hossain et al. We apply the Principal Components (abbreviated as PC) idea for constructing significant orthogonal components of indices in order to attain efficient frontiers that illustrate a true risk and return structure for a portfolio of stocks on the South African stock exchange under the SMI and ISMI formulations.

This study proposes a method of obtaining a smaller subset of orthogonal yet significant indices (via PC) that explain almost the equivalent volume of variation as using the entire basket of 8 indices. Once this common subset of indices has been established, theory and logic tells us that this newly constructed subset of indices retain the most information from the market as possible in the smallest dimension as possible. Hence the efficient frontiers constructed via this method should be the most accurate perception of risk and return that one can attain.

Furthermore, this empirical study proposes to estimate these new principal component frontiers for both the SMI and ISMI covariance structures. These PC frontiers will then be compared to the original SMI and ISMI frontiers, where the explanatory variables X_5 , X_6 and X_7 were used as the initial subset of indices. Finally this study will attempt to re- confirm from an earlier study, conducted in the Portfolio Theory chapter, that in fact the SMI model construes a less accurate perception of risk, where in fact it actually underestimates risk for portfolios with positively correlated residuals within the South African

stock market. In addition, this study will attempt to re-affirm through the new principal component frontiers that in fact the ISMI model in fact gives a more accurate perception of the true risk associated with a given level of return in the market in the presence of a positive dependence structure for portfolio covariances.

9.7.3 Methodology

In order to construct these principal component frontiers, the original basket of 8 indices is considered. The methodology of singular value decomposition is applied to this basket of indices to produce another basket of 8 orthogonal principal components. Next regressions are carried out on our portfolio of nine stocks in order to establish multiple index models, but with the newly acquired principal components. A common subset of significant orthogonal components is regressed with each of the nine stocks to produce our final models.

Principal component and mean variance analysis is straightforward as long as you have a way of computing the respective model betas, standard errors of regression (SER), stock means, covariance of indices, eigenvalues and eigenvectors of a covariance matrix. The procedure roughly followed in any time series or statistical package is

Step 1: Normalize data (in our case normalize the basket of indices) to mean zero and variance 1 and divide the data by the square root of the number of observations minus 1 in case a correlation matrix analysis is requested. Center data to have mean zero computing each variable minus its mean in case a covariance matrix is requested and divide the data by

the square root of the number of observations minus 1 in case a covariance matrix analysis is requested.

Step 2: Compute the singular value decomposition of the basket of indices. This results in the left and right singular vectors and the singular values of the data.

Step 3: Computation of component scores. This is obtained by multiplying the matrix of left singular vectors by the singular values.

Step 4: Computation of eigenvalues. This is computed squaring the singular values obtained in step 2.

Step 5: Computation of coefficients (eigenvectors). This is the matrix of right singular vectors obtained in step 2.

Step 6: Regress the first three significant principal components (PC) with each stock in the portfolio.

Step 7: Fit relative models of interest for each stock in portfolio i.e. *GARCH*(1, 1) (General Autoregressive Conditional Heteroskedastic), ARMA (Autoregressive Moving Average) models, State Space (SS) models and OLS (Ordinary Least Squares) model.

Step 8: Extract betas, SERs, stock means and respective model residuals for each estimation method

Step 9: Estimate respective covariance structures under SMI and ISMI frameworks

Step 10: Implement Sharpe Algorithm for efficient frontier construction in any programming language

Step 11: Plot Variance versus returns

9.7.4 Primary Findings

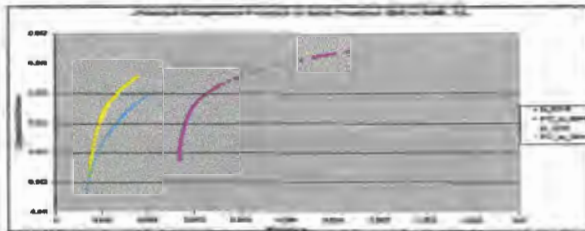


Fig 9.1: PC versus ISMI and SMI – OLS

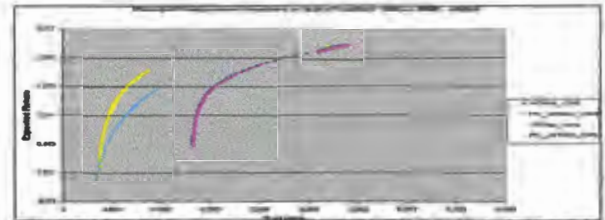


Fig 9.2: PC versus ISMI and SMI – ARMA

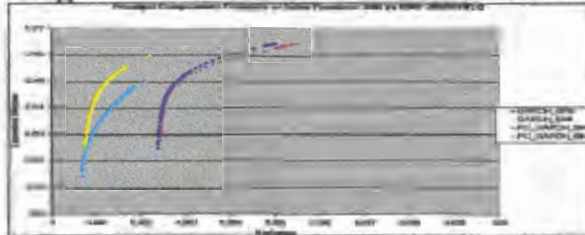


Fig 9.3: PC vs ISMI and SMI – GARCH

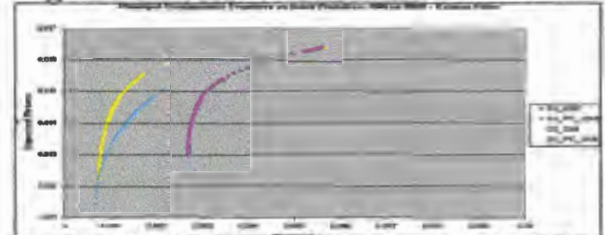


Fig 9.4: PC versus ISMI and SMI – SS

The above efficient frontiers in Figures 9.1, 9.2, 9.3 and 9.4 were constructed under four different estimation approaches: *OLS*, *ARMA*, *GARCH*(1, 1), and Kalman Filter under SMI, ISMI, PC-SMI, and PC-ISMI framework respectively.

Observing all four of the above illustrations, it is blatantly clear that a consistent empirical relationship persists to exist throughout the empirical investigation. In all 4 of the estimation techniques under examination it is primarily observed that the SMI frontier lies to the left of the ISMI frontier for all 4 estimation methods. Next calculating and plotting the new PC frontiers for both the SMI and ISMI methods for each estimation technique; it is perpetually established that the PC-ISMI frontier lies exactly on the original ISMI frontier for each of the 4 estimation techniques. Secondly the PC-SMI frontier consistently tends to lie to the right of the SMI frontiers for each of the estimation techniques under examination (Hossain (2006)).

9.7.5 Conclusion

Examining the findings from the previous section and considering the theory presented, it is very clear from the mathematical arguments disclosed that the principal components frontiers are the most accurate perception of the true risk and return relationship that can exist for any combination of stocks in an investor's portfolio. As a consequence of the results obtained in the previous section, it can be deduced from an empirical stand point that the ISMI and PC-ISMI frontiers are consistently equivalent whereas the PC-SMI and SMI frontiers are not in the four explored estimation techniques. Also it is established that the PC-SMI frontier lies consistently to the right of the SMI frontier.

As a result and consequence of these, it is empirically conclusive that the SMI estimation framework is consistently under estimating risk with regards to portfolios of stocks with a positively dependent covariance structure. The reasoning behind this is that the SMI covariance structure ignores the presence of covariance dependencies (i.e. off diagonal elements equal 0), thus the SMI efficient frontier will always lie to the left of the ISMI frontier in the case of portfolios exhibiting positively correlated residuals. In the case of portfolios demonstrating negative correlation, one should find that the SMI frontier will lie to the right of the ISMI frontier. In addition it is concluded that the original formulated ISMI multiple index model is in fact a more accurate risk evaluation method in the presence of positively correlated residuals. This was verified by a principal component frontier that was almost equivalent to the original ISMI in all four estimation methods (Hossain (2006)).

Chapter 10

Monte Carlo Simulation Study

This chapter will attempt to design dynamic simulated models for creating realistic sample paths of stock market index returns from which nine positively correlated stock returns are to be simulated via Monte Carlo methods. The stock market indices are simulated using Geometric Brownian motion models from which we construct a simulated portfolio of asset values. Adopting appropriate Monte Carlo methods the portfolio of stocks will be simulated exhibiting a $GARCH(1, 1)$ volatility process. We then compare the simulated portfolio under the Sharpe and Improved Sharpe formulations for the usual 4 estimation techniques, namely OLS, ARMA, State Space (Kalman Filter) and $GARCH(1, 1)$. An attempt will be made to show that the Sharpe formulation for the multiple index framework under estimates risk consistently for portfolios exhibiting strong positive correlation under a simulated environ.

10.1 Introduction

Modern Portfolio Theory is regarded as among 20th-century finance's most important analytical tools. Portfolio optimization was specifically created as a securities investment optimization tool. Given the expected return, and risk of a security, along with its correlation with other securities in the portfolio, a selection of stocks can be chosen such that the return is maximized for any given level of risk. Harry M. Markowitz, an economist at the University of Chicago, and William F. Sharpe at Stanford University, were awarded the

Nobel Prize in 1990 for their work over the previous 20 years in connection with Modern Portfolio Theory. Markowitz's original paper suggested that each portfolio of assets would have a given level of risk, and reward, but that for any level of risk, there was only one portfolio that would return an optimum reward. Conversely, for any level of reward, there would be only one portfolio that would minimize the risk.

Up to now, the research conducted thus far has empirically shown that the Sharpe Single Index (SSI) and Sharpe Multiple Index (SMI) formulations ignore the correlation structure of residuals of financial portfolios. To address this problem, Hossain et al has introduced the assumption of correlated residuals into the model (i.e. the Improved Sharpe Multiple Index Model (ISMI)), which gives rise to an entirely new covariance structure with non zero off diagonal elements. It was shown from previous empirical studies, conducted on the South African stock exchange, that for portfolios with positively correlated assets, the Sharpe formulation under estimates the true risk reward structure under both the multiple and single index environs. This was empirically demonstrated for 4 estimation methods namely, *GARCH*(1, 1), State Space models, *OLS* and *ARMA*.

These empirical claims were further backed up by a principal components' study conducted in the previous chapter. The discussion that follows presents a Monte Carlo simulation that will attempt to further back up our empirical discovery that the Sharpe formulation under estimates risk (efficient frontier to the left) for portfolios with positively correlated residuals. To further prove this empirical result we shall perform a simulated stock market study by means of employing Monte Carlo methods.

This chapter will attempt to design dynamic simulated models for creating realistic sample paths of stock market index returns from which nine stock returns are to be simulated via Monte Carlo methods. An attempt will be made to show that the Sharpe Multiple Index formulation under estimates risk for simulated financial portfolios exhibiting strong positive correlation. This chapter is organised as follows. Section 2 describes the Geometric Brownian motion model, which is the model we shall employ to simulate our stock market indices. Section 3 develops the framework for the Monte Carlo simulation of our Geometric Brownian motion model and the techniques applied to simulate a nine stock portfolio. Section 4 highlights the step by step methodology adopted for executing the simulation at hand. Section 4 describes the empirical results obtained, from applying the Sharpe Multiple Index (SMI) and the Improved Sharpe Multiple Index (ISMI) formulations on the simulated data. Section 5 provides concluding comments.

10.2 Brownian Motion

10.2.1 Introduction

A stochastic process is a system which evolves in time while undergoing chance fluctuations. The time can change discretely or continuously. The variable can have discrete values or continuous values. The binomial model assumed a discrete-valued and discrete-time stochastic process. In Financial Mathematics, the stochastic behaviour of share prices is studied under the Markovian approximation. A Markovian process is a process whose future does not depend on its past history. The prediction of a future event depends only on

the present state for a Markovian process. The Markov process is related to the weak form of the efficient-market hypothesis.

The observation that, when suspended in water, small pollen grains are found to be in a very animated and irregular state of motion, was first systematically investigated by Robert Brown in 1827, and the observed phenomenon took the name Brownian motion because of his fundamental pioneering work. Brownian motion, which is a limiting process of a random walk, is a Markov process with a continuous state space and a continuous time set. The process and its many generalisations occupy a central role in an option pricing model. We can derive the diffusion equation underlying the Brownian motion process.

10.2.2 Einstein's Brownian Motion

Einstein mathematically formulated the notion of Brownian Motion. The following equation

$$\frac{\partial u}{\partial t} = \frac{\sigma^2}{2} \frac{\partial^2 u}{\partial x^2} \quad (1)$$

was used by him to solve the probability $u(x, t)$ of particles being found in x at t for one-dimensional Brownian motion. In equation (1), σ^2 is regarded as the diffusion coefficient.

The formal solution of (1) is a Gaussian function or a normal function

$$u(x, t) = \frac{1}{\sqrt{2\pi\sigma^2 t}} e^{-x^2/2\sigma^2 t} \quad (2)$$

where

$$E[X] = 0 \text{ and } Var[X] = \sigma^2 t \quad (3)$$

The expectation and variance of x are found as

$$\begin{aligned} E[x] &= \int_{-\infty}^{\infty} xu(x)dx \\ var[x] &= \int_{-\infty}^{\infty} x^2u(x)dx = E[x]^2 \end{aligned}$$

The above definition for the expectation and variance of x is considered when the probability function $u(x)$ is continuous.

10.2.3 Random Walk

We consider one-dimensional discrete random walk which yields Brownian motion under the continuum limit. Note that when the mean displacement is not zero, the system is not the Brownian motion by standard definition but we will conventionally call it a Brownian motion. Let $\{X_i\}$ be random variables with

$$P(x_i = k) = p \text{ and } P(x_i = -k) = q \quad (4)$$

where k is the size of the i - th step, with probability p that the process moves up in a positive direction and with probability q that it moves downward to a negative direction. It is easily shown that the expected values and variance of (4) are

$$E(x_i) = (p - q)k \text{ and } var(x_i) = 4pqk^2 \quad (5)$$

Let X_n denote the position of a random walk process after n steps on an X_n plane. Let us consider X_n and define it such that $X_k = x_1 + x_2 + \dots + x_n$, for $k = 1, 2, \dots, n$. The stochastic process $\{X_n, n \geq 0\}$ is called a random walk process. The mean and variance of X_n are therefore

$$E(X_n) = (p - q)nk \text{ and } var(X_n) = 4pqk^2n \quad (6)$$

Recall the definition for Taylor expansion from first year mathematics: In the vicinity of a point x_0 , the value of a function $f(x)$ is Taylor expanded as follows:

$$f(x) = f(x_0) + (x - x_0) \left[\frac{\partial f(x)}{\partial x} \right]_{x=x_0} + \frac{1}{2}(x - x_0)^2 \left[\frac{\partial^2 f(x)}{\partial x^2} \right]_{x=x_0} + \dots \quad (14)$$

The Taylor expansion of equation (13) the follows as:

$$\begin{aligned} & u(x, t) + \lambda \frac{\partial u(x, t)}{\partial t} + O(\lambda^2) \\ = & u(x, t) + k(q - o) \frac{\partial u(x, t)}{\partial x} + \frac{k^2}{2} \frac{\partial^2 u(x, t)}{\partial x^2} + O(K^3) \\ \Rightarrow & \frac{\partial u(x, t)}{\partial t} = \left[(q - p) \frac{k}{\lambda} \right] \frac{\partial u(x, t)}{\partial x} + \frac{k^2}{2\lambda} \frac{\partial^2 u(x, t)}{\partial x^2} \end{aligned} \quad (15)$$

under the assumption $\lambda, k \rightarrow 0$. With use of equation (11) and (15) we obtain the partial differential equation

$$\frac{\partial u(x, t)}{\partial t} = -\mu \frac{\partial u(x, t)}{\partial x} + \frac{\sigma^2}{2} \frac{\partial^2 u(x, t)}{\partial x^2} \quad (16)$$

This is known as the *forward Kolmogorov equation* for the drift rate μ and the diffusion rate σ^2 . The formal solution to (16) is

$$u(x, t) = \frac{1}{\sqrt{2\pi\sigma^2 t}} \exp \left[-\frac{(x - \mu t)^2}{2\sigma^2 t} \right] \quad (17)$$

This again a Normal function with peak at $x = \mu t$ and spread $\sigma^2 t$. Note that these attributes are time dependent, hence they grow as time moves forward. From equation (11) we find that $\Delta X = \sigma \sqrt{\Delta t}$. Without drift, when Δt goes to 0,

$$(dX(t))^2 = \sigma^2 dt \quad (18)$$

thus the generalised Brownian motion with drift is written as

$$dX(t) = \mu dt + \sigma dZ \quad (19)$$

with the stochastic property that $(dZ)^2 = dt$.

10.2.4 Geometric Brownian Motion

When $X(t)$ is a Brownian Motion with drift rate μ and diffusion rate σ^2 , the stochastic process defined by

$$y(t) = e^{x(t)} \quad (20)$$

is called the *Geometric Brownian Motion*. The probability density function for $y(t)$ is

$$u(y, t) = \frac{1}{y\sigma\sqrt{2\pi t}} \exp \left[-\frac{(\ln y - \mu t)^2}{2\sigma^2 t} \right], y > 0 \quad (21)$$

The mean and variance of $y(t)$ are, respectively,

$$\begin{aligned} E[y(t)|y(0) = y_0] &= y_0 \exp \left(\mu t + \frac{\sigma^2 t}{2} \right) \\ \text{var}[y(t)|y(0) = y_0] &= y_0^2 e^{2\mu t} (e^{\sigma^2 t} - 1) \end{aligned} \quad (22)$$

10.2.5 Ito's Lemma

A prominent generalisation of Brownian motion processes is the class of processes known as Ito processes. Since publication of the seminal paper of Black and Scholes in 1973, Ito processes have remained in the centre stage of continuous-time finance. If X follows an Ito process:

$$dX(t) = a(X, t)dt + b(X, t)dZ(t) \quad (23)$$

where the parameters a and b are functions of the value of the underlying variable, X , and t and $dZ(t)$ is a Wiener process. Ito's lemma shows that a function, Y , of X and t follows the process

$$dY = \left(\frac{\partial Y}{\partial X} a + \frac{\partial Y}{\partial t} + \frac{1}{2} \frac{\partial^2 Y}{\partial X^2} b^2 \right) dt + \frac{\partial Y}{\partial X} b dZ \quad (24)$$

Thus Y also follows an Ito process with drift rate

$$\frac{\partial Y}{\partial X}a + \frac{\partial Y}{\partial t} + \frac{1}{2} \frac{\partial^2 Y}{\partial X^2} b^2 \quad (25)$$

and variance rate

$$\left(\frac{\partial Y}{\partial X} \right)^2 b^2 \quad (26)$$

Proof

By the Taylor expansion of ΔY , which is a function of X and t , we obtain

$$\begin{aligned} \Delta Y = & \frac{\partial Y}{\partial X} \Delta X + \frac{\partial Y}{\partial t} \Delta t + \frac{1}{2} \left(\frac{\partial^2 Y}{\partial X^2} \Delta X^2 + \frac{\partial^2 Y}{\partial X \partial t} \Delta X \Delta t + \frac{\partial^2 Y}{\partial t^2} \Delta t^2 \right) \\ & + \frac{1}{3!} \left(\Delta X \frac{\partial}{\partial X} + \Delta t \frac{\partial}{\partial t} \right)^3 Y + \dots \end{aligned} \quad (27)$$

The discretised form of equation (23) is

$$\Delta X = a(X, t) \Delta t + b(X, t) \sqrt{\Delta t} \quad (28)$$

In the limit of $\Delta t \rightarrow 0$, we realise that $\Delta X^2 \approx b^2(X, t) \Delta t$. Substituting this into (27) we obtain

$$\begin{aligned} dY = & \frac{\partial Y}{\partial X} dX + \frac{\partial Y}{\partial t} dt + \frac{1}{2} \frac{\partial^2 Y}{\partial X^2} b^2(X, t) dt \\ = & \left(\frac{\partial Y}{\partial X} a + \frac{\partial Y}{\partial t} + \frac{1}{2} \frac{\partial^2 Y}{\partial X^2} b^2 \right) dt + \frac{\partial Y}{\partial X} b dZ \end{aligned} \quad (29)$$

The second line in the right-hand side has been obtained by using equation (23).

10.2.6 Application to Geometric Brownian Motion

Let us assume the following Brownian motion

$$dM = r dt + \sigma dZ \quad (30)$$

Then the governing equation for the variable $S = e^M$ is

$$\begin{aligned} dS &= S \left(r + \frac{\sigma^2}{2} \right) dt + S\sigma dZ \\ \Rightarrow \frac{dS}{S} &= \left(r + \frac{\sigma^2}{2} \right) dt + \sigma dZ \end{aligned} \quad (31)$$

which has been obtained using Ito's lemma and

$$\frac{\partial S}{\partial M} = e^M = S, \quad \frac{\partial S}{\partial t} = 0, \quad \frac{\partial^2 S}{\partial M^2} = e^M = S \quad (32)$$

If there is no uncertainty in the stock market, when the expected rate of return on the stock is μ , the stock price at time T is $S_t e^{\mu(T-t)}$ which is a solution of the equation

$$\frac{dS}{S} = \mu dt \quad (33)$$

If the uncertainty of the stock price is described by the geometric Brownian motion,

$$\begin{aligned} \frac{dS}{S} &= \mu dt + \sigma dZ \\ \Rightarrow dS &= \mu S dt + \sigma S dZ \end{aligned} \quad (34)$$

where σ is the market volatility. Comparing (32) with (34) we can see that the stock price undergoes geometric Brownian motion. We have now validated the theoretical reasoning behind using a Geometric Brownian Motion model for our stock price processes. Next we develop the framework for the Monte Carlo simulation of our Geometric Brownian motion model for stock indices.

10.3 Monte Carlo Simulation of Geometric Brownian Motion

In this section we present the steps to perform the simulation of the main stochastic processes used in the portfolio optimization applications. Consider that the price P of a commodity

follows a Geometric Brownian Motion, which is given by the following stochastic equation:

$$dP = \alpha P dt + \sigma P dZ \quad (1)$$

Where $dz = \text{Wiener increment} = \varepsilon dt^{0.5}$, where ε is the standard normal distribution; α is the drift (or capital gain rate); and σ is the volatility of P . By using the equation of *total investment return* $\mu = \alpha + \delta$, where μ is also the risk adjusted discount rate for P ; and δ is the dividend yield (or convenience yield in case of commodities). We can rewrite the stochastic equation as:

$$dP = (\mu - \delta) P dt + \sigma P dZ \quad (2)$$

For the risk-neutral version of this equation, just replace the risk-adjusted discount rate μ by the risk-free interest rate r to obtain the risk-neutral stochastic equation:

$$dP = (r - \delta) P dt + \sigma P dZ \quad (3)$$

Using a logarithm transformation and applying the Itô's Lemma, we can reach the equations for the prices simulation. For this study we are going to use the 'real' simulation of a Geometric Brownian Motion (GBM) and ignore the risk-neutral case. The real simulation of a GBM uses the real drift α . The price P_t at the future instant t is given by:

$$P_t = P_0 \exp \left\{ \left(\alpha - \frac{1}{2} \sigma^2 \right) \Delta t + \sigma \varepsilon_t \sqrt{\Delta t} \right\} \quad (4)$$

where ε is the standard normal distribution, i.e. $\varepsilon \sim N(0, 1)$. The simulation of the real prices using the above equation is done by sampling the standard Normal distribution $N(0, 1)$ and obtaining the correspondent values for P_t . These values of P can be used to calculate the (real) values of the portfolio V . Remember, with the real drift simula-

tion of the underlying asset P , the required discount rate is the risk-adjusted one μ . One important feature of the above discrete-time equations is that the discretization from the continuous-time model is exact. In other words, you don't need to use small time increments Δt in order to get a good approximation. You can use any Δt that the simulation equation is valid.

10.4 The Methodology

The previous section provides the tools that are used in the simulation of an appropriate stock market index. For this study we shall simulate 3 stock market indices employing such a tool which we shall represent as index1, index2 and index 3 respectively. The price process of these 3 indices shall be generated using equation (3) from the previous section, recall:

$$P_t = P_0 \exp \left\{ \left(\alpha - \frac{1}{2} \sigma^2 \right) \Delta t + \sigma \varepsilon_t \sqrt{\Delta t} \right\} \quad (1)$$

We simulate 200 observations for each index, for $i = 1, 2, 3$. We set $\Delta t = 0.05$, $\mu = 1$, $\sigma = 1$ and set $P_0 = 50$. Figures 1 to 3 below represent the price processes for each of the three stock market indices simulated.

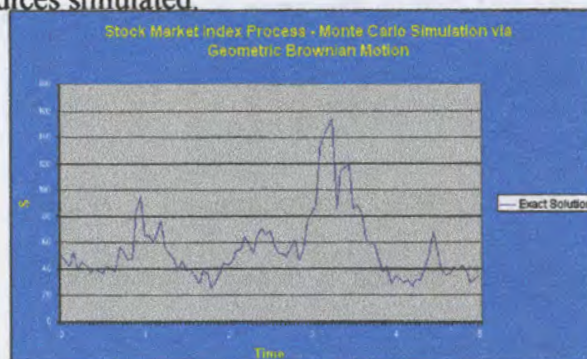


Figure 10.1: Simulation of first Index

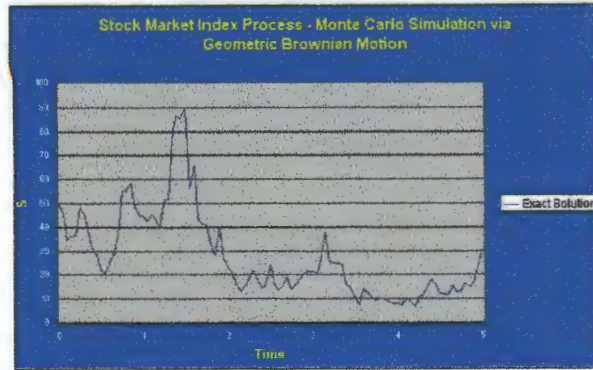


Figure 10.2: Simulation of Second Index

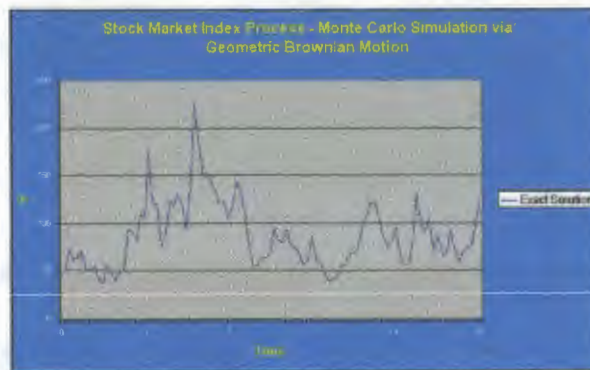


Figure 10.3: Simulation of thrid Index

The next step is to make our data stationary by means of calculating log returns for each index. Now we are ready to construct a nine stock portfolio using the simulated indices.

For each stock, R_{it} , $i = 1, \dots, 9$, the following regression model is used:

$$R_{it} = \beta_0 + \beta_1 \text{Index1}_t + \beta_2 \text{Index2}_t + \beta_3 \text{Index3}_t + e_t, \quad i = 1, \dots, 9 \quad (2)$$

The model parameters, $\beta_0, \beta_1, \beta_2$ and β_3 are varied against actual market benchmarks used from past studies. We vary them so as to obtain a positively correlated portfolio of stocks.

The crucial step that follows is the simulation of each e_t . We model e_t in the following

manner

$$e_t = \varepsilon_t \sigma_t \quad (3)$$

where

$$\sigma_t^2 = \alpha_0 + \alpha_1 \sigma_{t-1}^2 + \phi_1 e_{t-1}^2 \text{ and } \varepsilon_t \sim N(0, 1) \quad (4)$$

Equation (4) is our elementary *GARCH*(1, 1) process. In order to kickstart the simulation of the above stock return process, an initial estimate of σ_{1i} is required for $i = 1, \dots, 9$. Such an estimate is acquired from residual values from previous multiple index models of similar structure. The rest of the simulation then follows recursively from there on until the 200th return is calculated for each of the nine stocks. Finally we have our simulated portfolio of nine stocks. The next phase of the simulation is the model building phase. We highlight the steps as follows:

Step 1: Regress the three simulated indices with each stock in the simulated portfolio.

Step 2: Fit relative models of interest for each stock in portfolio i.e. *GARCH*(1, 1) (General Autoregressive Conditional Heteroskedastic), ARMA (Autoregressive Moving Average) models, State Space (SS) models and OLS (Ordinary Least Squares) model.

Step 3: Extract betas, Standard Errors of Regression (SER), stock means and respective model residuals for each estimation method i.e. *GARCH*(1, 1), ARMA, OLS and Kalman Filter (SS)

Step 4: Estimate the respective covariance structures for each estimation technique under SMI and ISMI frameworks

Step 5: Implement Sharpe Algorithm for efficient frontier construction in any programming language

Step 6: Plot standard errors versus returns

The next section illustrates the results from applying the above steps to our simulated data

10.5 Empirical Study

Applying the above methodology we obtain the following efficient frontiers.

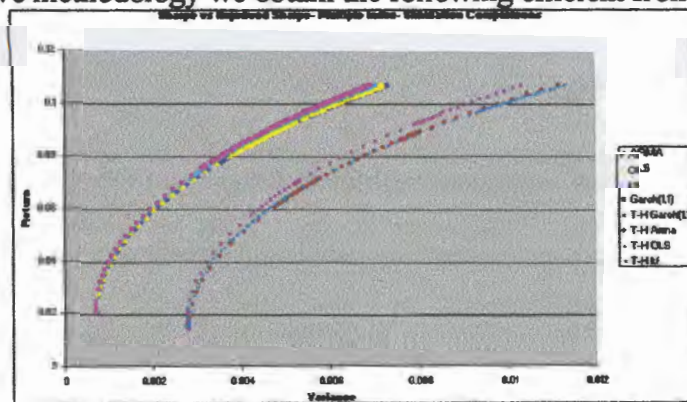


Fig 10.4: SMI vs ISMI: Simulation Results

Immediately a consistent relationship can be ascertained from figure 4; that is efficient frontiers estimated under the SMI formulation lie consistently to the left of the ISMI framework. In a financial context, this is interpreted as follows that for a given level of return, the SMI formulation is implying a lower level of risk in contrast to the ISMI formulation for portfolios with positively correlated residuals. From an investment perspective the SMI formulation is deemed to be superior from a risk point of view. Retrospectively the SMI formulation has off-diagonal elements as zero it thus ignores the correlation structure of the residuals between stocks and therefore ignores the risk associated with such a structure. Hence, the SMI formulation gives the investor a false perspective as to the actual level of

risk carried by an investment portfolio in the case of a correlated residual structure,. this can be dangerous especially for risk-averse investors. In contrast the ISMI framework incorporates such an assumption into its formulation and therefore adjusts the risk-return level appropriately.

Furthermore, it is simultaneously established that in both formulations the $GARCH(1, 1)$ methodology has performed superior in contrast to the other 3 methods (Hossain (2006)).

10.5.1 Conclusion

In conclusion, upon examination of the findings from the previous section and considering the theory presented, it is clear that the SMI formulation is a less accurate estimator of a portfolio's risk-return frontier in the presence of correlated residuals. Furthermore with respect to the case of positively correlated residuals, we find that the SMI formulation under-estimates risk consistently in contrast to ISMI (Hossain(2006)). This conclusion was further confirmed by our market and principal components studies conducted in previous chapters. Finally to reconfirm the central hypothesis, the $GARCH(1, 1)$ methodology has consistently proven to be the superior methodology in both formulations as shown in previous studies ..

Chapter 11

Summary And Directions For Further Research

This thesis contributed new model formulations and important empirical discoveries in the area of modern portfolio theory and portfolio optimisation. This Chapter summarizes the contributions made in the thesis, and also points out directions for further research.

Chapters 2,3,4,5 and 6 were devoted towards the introduction and background of the 4 modelling techniques that were to be utilised in comparative studies in forthcoming chapters regarding portfolio optimisation and portfolio models. Chapter 2 introduces some basic and fundamental stock market theory and served as a foundation for ensuing chapters. Chapter 3 introduced the OLS (Ordinary Least Squares) framework, chapter 4 the *ARMA* models, chapter 5 presented State Space models and finally *GARCH* models and its extensions were introduced in chapter 6. These 4 chapters dealt with the estimation procedures and the model building processes involved of these 4 modelling paradigms

The Sharpe algorithm for Efficient frontier construction and the optimisation process involved is one of the initial concepts introduced in Chapter 7. The Chapter considered multiple and single index portfolio models of Sharpe. One of the crucial contributions of this chapter is the empirical discovery of the *GARCH*(1, 1) model's dominating behaviour in a risk return setting for both the Sharpe multiple and single index settings. The dominance of *GARCH*(1, 1) is also established amongst its extensions within the volatility modelling paradigm for the Sharpe single index case.

The major contribution of Chapter 8, and this thesis, is a new general portfolio model formulation for stocks, namely the Improved Sharpe formulation proposed by Troskie and Hossain for both the single and multiple index settings. This chapter demonstrated that for stock portfolios that illustrate strong positive correlation between residuals it was found that the Sharpe formulation under-estimates risk in both the multiple and single index cases in comparison to the Improved Sharpe. Conversely for the negative correlation case this chapter demonstrated that the Sharpe formulation over-estimates risk when compared with the Improved Sharpe for both the single and multiple index cases. Finally this chapter illustrated that the Improved Sharpe model proposed by this thesis is a more accurate perception of the risk-return relationship of a portfolio of stocks. The Chapter also dealt with an empirical investigation into the domineering behaviour of the $GARCH(1, 1)$ model in a risk-return setting for both the Sharpe and Improved Sharpe models for both the multiple and single index scenarios as well

A unique application of PCA (principal Components Analysis) was considered in Chapter 9, where we applied the Principal Components (abbreviated as PC) idea for constructing significant orthogonal components of indices in order to attain efficient frontiers that illustrate a true risk and return structure for a portfolio of stocks on the South African stock exchange under the Sharpe Multiple Index (SMI) and Improved Sharpe Multiple Index (ISMI) formulations. The empirical evidence illustrated that the application of PC in efficient frontier construction supports the risk-return structure estimated under the ISMI formulation. Furthermore the empirical results in this chapter supports the evidence from chapter 8 that the SMI model gives an inaccurate risk structure of an investor's financial

portfolio and therefore is consistently under-estimating risk in portfolios with positively correlated residuals.

Finally Chapter 10, presented a Monte Carlo simulation study, where appropriate Monte-Carlo methods were adopted and the portfolio of stocks was simulated exhibiting a $GARCH(1,1)$ volatility process. The behaviour of the simulated portfolio was compared under two developed multiple index portfolio frameworks, the SMI and the ISMI portfolio models proposed by Hossain et al. It was shown that the SMI formulation for the multiple index framework consistently under-estimates risk for portfolios exhibiting strong positive correlation under a Monte-Carlo simulated environ and explore the consistency of this claim under four estimation methods namely, $GARCH$, $ARMA$, OLS and SS (Kalman Filter) models. The Monte Carlo simulation study serves as further evidence that supports claims from the PC study and finally supports the empirical evidence and claims presented in chapter 8 regarding the SMI and ISMI models.

It should be clear that this thesis covered a number of interesting topics in, model selection, model building, portfolio optimisation, modern portfolio theory, principal components analysis and Monte Carlo simulation. It is expected that the Improved Sharpe model proposed by this thesis will become a preferred portfolio model somewhere in the not too distant future. Many other research topics were identified during the course of this study, and offers scope for further timely research to be conducted in modern portfolio theory.

References

- Afleck-Graves J.F., Money A.H. and Troskie C.G., (1979), "A principal component index subject to constraints", *The Investments Analysts Journal*, **14**, 45-52.
- Akaike H., (1969), "Fitting autoregressive models for Prediction", *Annals of the Institute of Statistical Mathematics*, **21**, 243-247.
- Akaike H., (1973) "Information theory as an extension of the maximum likelihood principle." pages 267-281 in *Second international symposium on information theory*. B.N. Petrov and F Csaki (editors). Akademiai Kiado, Budapest.
- Anderson T.W., (1958), '*An Introduction to Multivariate Statistical Analysis*', New York, John Wiley and Sons, Inc, London.
- Attanasio, O.P., (1991), "Risk, time-varying second moments and market efficiency", *Review of Economic Studies*, **58**, 479-94.
- Attanasio, O.P. and S. Wadhvani, (1989), "Risk: Gordon's Growth model and the predictability of stock market returns", Stanford Center for Economic Policy Research Discussion Paper Series:161,47.
- Baillie, R. T., and T. Bollerslev (1989a), "Common stochastic trends in a system of real exchange rates", *Journal of Finance*, **44**, 1676-181.
- Baillie, R., and R. DeGennaro, (1990), "Stock returns and volatility", *Journal of Financial and Quantitative Analysis*, **25**, 203-214.
- Barberis, N., A. Shleifer, and R. Vishny, (1998), "A model of investor sentiment", *Journal of Financial Economics*, **49**, 307-343.
- Barrett, B. E. and Gray, J. B. (1994), "A Computational Framework for Variable Selection in Multivariate Regression", *Statistics and Computing*, **4**, 203-212.
- Bedrick, E.J., and C.-L. Tsai (1994) "Model selection for multivariate regression in small samples", *Biometrics*, **50**, 226-231.
- Berndt, E., H. R. Hall, B.H., and J. Hausman, (1974), "Estimation and Inference in Non-linear Structural Models", *Annals of Economic and Social Measurement*.

- Bilodeau, M. and Brenner, D., (1999), *Theory of Multivariate Statistics*, Springer-Verlag New York, Inc., New York.
- Black, F., (1976), "Studies of Stock Price Volatility Changes", *Proceedings of the 1976 Meetings of the American Statistical Association, Business and Economical Statistics Section*, 177-181.
- Black, F. and Litterman R. (1991), "Global asset allocation with equities, bonds and currencies", *Review of Financial Studies*, 4:2, 315-342.
- Bollerslev T. (1986), "Generalized autoregressive conditional heteroskedasticity", *Journal of Econometrics*, 31, 307-327.
- Bollerslev T. (1987), "A conditional heteroskedastic model for speculating prices and rates of return", *Review of Economics and Statistics*, 69, 542-547.
- Bollerslev, T., Engle, R.F. and Wooldridge, J.M. (1988), "A Capital Asset Pricing Model with Time-varying Covariances", *Journal of Political Economy*, 96, 116-131.
- Bollerslev, T., and J. Wooldridge (1992), "Quasi-maximum Likelihood Estimation and Inference in Dynamic Models with Time-varying Covariances", *Econometric Reviews*, 11, 143-172.
- Box, G.E.P. and Jenkins, G.M. (1970), *Time series analysis: forecasting and control*, San Francisco: Holden-Day.
- Box, G.E.P., Jenkins, G.M., and Reinsel, G.C. (1994), *Time series analysis: forecasting and control* 3rd ed. Prentice Hall, Englewood Cliffs, N.J.
- Box, G.E.P., and Tiao, G.C., (1973). *Bayesian inference in statistical analysis*,. John Wiley, New York
- Breedon, D. T. (1979), "An intertemporal asset pricing model with stochastic consumption and investment opportunities", *Journal of Financial Economics*, 7, 265-296.
- Breiman, L. and Friedman, J.H., (1997), "Predicting Multivariate Responses in Multiple Linear Regression", *Journal of the Royal Statistical Society*, 59 (No. 1), 3-54.
- Brockwell P.J. and Davis RA (1996), *Introduction to Time Series and Forecasting*, New York: Springer-Verlag.
- Broomhead, D.S. and King, G.P. (1986), "Extracting Qualitative Dynamics from Experimental Data", *Physica D*, 20, 217-236.

- Buchstaber, V.M. (1994), "Time Series Analysis and Grassmannians", *Amer. Math. Soc. Transl.*, **162**, 1-17.
- Burnham, K.P. and Anderson, D.R. (1998), "*Model Selection and Inference*", New York: Springer-Verlag.
- Burnham, K.P. & Anderson, D.R. (2002), "*Model Selection and Multimodel Inference: A Practical Information-Theoretic Approach*", 2nd edn. Springer-Verlag, New York, NY.
- Buser, S.A., (1977), "Mean-variance portfolio selection with either a singular or non-singular variance-covariance matrix", *Journal of Financial and Quantitative Analysis*, 347-361
- Campbell, J. (1987), "Stock Returns and the Term Structure", *Journal of Financial Economics*, **18**, 373-399.
- Chamberlain, G., and M. Rothschild, (1983), "Arbitrage, Factor Structure, and Mean-Variance Analysis on Large Asset Markets", *Econometrica*, **51**, 1281-1304.
- Chatfield, C. (1994), "*The Analysis of Time Series An Introduction*", London:Chapman and Hall.
- Christie, A.A., (1982), "The Stochastic Behavior of Common Stock Variances - Value, Leverage and Interest Rate Effects", *Journal of Financial Economics*, **3**, 145-166
- Clark, A.E., (2003), 'Model selection-regression and time series applications', University of Cape Town, Unpublished MSc Thesis
- Cowles A., (1933), "Can Stock Market Forecasters Forecast?", *Econometrica*, **1**, 309-24.
- Cowles A., (1944), "Stock Market Forecasting", *Econometrica*, **12**, 206-14
- Cowles A., (1960), "A Revision of Previous Conclusions Regarding Stock Price Behaviour", *Econometrica*, **28**, 909-15
- Cowles A. and Jones H.E., (1937), "Some A Posteriori Probabilities in Stock Market Action", *Econometrica*, **5**, 280-94
- Daniel, K., D. Hirshleifer, and A. Subrahmanyam, (1998), "Investor psychology and security market under- and overreaction", *Journal of Finance*, **53**, 1839-1885.
- Danilov, D.L., (1997), "Principal Components in Time Series Forecast", *Journal of Computational and Graphical Statistics*, **6**, 112-121.

- DeLong, J.B., A. Shleifer, L.H. Summers, and R.J. Waldmann, (1990), "Positive feedback investment strategies and destabilizing rational speculators", *Journal of Finance*, **45**, 379-395.
- Diebold, F., and M. Nerlove (1989), "The dynamics of exchange rate volatility: a multivariate latent factor ARCH model", *Journal of Applied Econometrics*, **4**, 1-21.
- Ding, Z., Granger, C. W. J. and Engle, R. F., (1993), "A long memory property of stock market returns and a new model", *Journal of Empirical Finance*, **1**, 83-106.
- Draper N.R., (1966), "*Applied Regression Analysis*", Wiley.
- Elton, E.J., Gruber, M.J. and Padberg, M.W. (1977), "Simple rules for optimal portfolio selection: the multi group case", *Journal of Financial and Quantitative Analysis*, 329-345
- Engle, R.F. (1982), "Autoregressive Conditional Heteroscedasticity with Estimates of the Variance of UK Inflation", *Econometrica* **50**, 987-1008.
- Engle, R.F. and G. Gonzalez-Rivera(1991), "Semi-Parametric ARCH Models," *Journal of Business and Economic Statistics*, **9**, 345-359.
- Engle, R.F., T. Ito and Wen-Ling Lin. (1990), "Meteor Showers or Heat Waves? Heteroskedastic Intra-Daily Volatility in the Foreign Exchange Market", *Econometrica*, May, **58:3**, pp.525-42.
- Engle, R. F. and A. P. Rodrigues (1989), "Tests of international CAPM with time-varying covariances", *Journal of Applied Econometrics*, **4**, pp. 119-138.
- Fama, E., (1970), "Efficient capital markets: A review of theory and empirical work", *Journal of Finance*, **25**, 383-417.
- Fama, E., (1991), "Efficient capital markets: II", *Journal of Finance*, **46**, 1575-1617.
- Fama, E. F., Fisher, L., Jensen, M., & Roll, R., (1969), "The adjustment of stock prices to new information", *International Economic Review*, **10**, 1-21.
- Fama, E. and French, K., (1988), "Permanent and temporary components of stock prices", *Journal of Political Economy*, **96**, 246-273.
- Fama, E. and MacBeth (1973), "Risk, Return, and Equilibrium: Empirical Tests", *Journal of Political Economy*, **81**, 607-36

- Fernandez, C., and Steel, M.F.J. (1998), "On Bayesian Modelling of Fat Tails and Skewness", *Journal of the American Statistical Association*, **93**, 359-371.
- Fowler, A.C. and Kember, G., (1998), "Singular system analysis as a moving-window spectral method", *European Journal of Applied Mathematics*, **9**, 55-79.
- French, Schwert and Stambaugh (1987), "Expected Stock Returns and Volatility", *Journal of Financial Economics*, **19**, 3-29.
- Gallant, A.R. and D.W. Nychka (1987), "Seminonparametric maximum likelihood estimation", *Econometrica*, **55**, 363-90.
- Gallant, A. Ronald and George Tauchen, (1989), "Seminonparametric estimation of conditionally constrained heterogeneous processes: Asset pricing applications", *Econometrica*, **57**, 1091-1120.
- Gallant, A. R., P. E. Rossi, and G. Tauchen, (1992), "Stock Prices and Volume", *Review of Financial Studies*, **5**, 199-242.
- Gill, P.E, Murray, W., Wright, M.H., (1991), *Numerical Linear Algebra and Optimization*, California: Addison-Wesley Publishing Company.
- Glosten, Lawrence R., Ravi Jagannathan, and David E. Runkle, (1993), "On the relation between the expected value and the volatility of the nominal excess return on stocks", *Journal of Finance*, **48**, 1779-1801.
- Goldberger, A.S., (1962), "Best linear unbiased prediction in the generalized linear regression model", *Journal of the American Statistical Association*, **5**, 369-375.
- Golyandina, N., Nekrutkin, V., Zhigljavsky, A., (2001), "Analysis of Time Series Structure SSA and Related Techniques", *Monographs on Statistics and applied Probability*, **90**, Chapman and Hall/CRC.
- Granger, C. W. J. and Andersen, A. P.: (1978), *Introduction to Bilinear Time Series Models*, Vandenhoeck and Ruprecht, Göttingen.
- Hamilton, J. D. (1994), *Time Series Analysis*, Princeton Un. Press.
- Harvey, A.C., (1981), *The econometric analysis of time series*, Oxford
- Harvey A.C., (1989), *Forecasting, Structural Time Series Models, and the Kalman Filter*, Cambridge University Press, Cambridge.

- Harvey, A.C., (1993), "*Time Series Models*", Prentice Hall International.
- Hossain N., Troskie C.G. and Guo R. (2005), "GARCH(1,1) Volatility Modeling in Modern Portfolio Optimization", *Proceedings of the International Workshop on Recent Advances Stochastic Operations Research*, Canmore, Alberta, Canada, August 25-26, 2005, pp 80-87. (ISBN 4-939146-03-0). Edited by Tadashi Dohi, Shunji Osaki and Katsushige Sawaki
- Hossain N., Troskie C.G. and Guo R. (2005), "Comparisons of the Ex Post Efficient Portfolios under Garch(1,1) Modeling and Garch Model Extensions.", *Proceedings of 10th International Conference on Industrial Engineering, Theory, Applications and Practice*, Clearwater, Florida, United States of America, December 4-7, 2005, pp 419-425 . ISSN: 0-9654506-1-9, Editors: TK Fredericks, SE Butt, A. Pennathur, JE Fernandez, A. Mital.
- Hossain N., Troskie C.G. and Guo R. (2005), "Comparisons of the Ex Post Efficient Portfolios Generated from Sharpe and Troskie-Hossain Index Models.", *Proceedings of 10th International Conference on Industrial Engineering, Theory, Applications and Practice*, Clearwater, Florida, United States of America, December 4-7, 2005, pp 412-417 . ISSN: 0-9654506-1-9, Editors: TK Fredericks, SE Butt, A. Pennathur, JE Fernandez, A. Mital.
- Hossain N., Troskie C.G. and Guo R. (2006), "An Application of Principal Components Analysis to Portfolio Optimization under the Sharpe Multiple Index Framework.", *Proceedings of International Workshop of Financial Engineers*, Hong Kong, June 20-22, 2006, pp 396-401. (ISBN-13: 978-988-98671-3-3). Edited by S.I. Ao, J.A. Lee, O. Casitllo, P. Chaudhuri, D.D. Feng
- Hossain N., Troskie C. and Guo R. (2006), "Comparisons of Ex Post Efficient Portfolios Generated from Sharpe and Improved Sharpe Multiple Index Models; An Empirical Study concerning the Negative Correlated Case", *Proceedings of 11th International Conference on Industrial Engineering, Theory, Applications and Practice*, Nagoya, Japan, October 24-27, 2006. (Accepted)
- Hossain N., Troskie C. and Guo R. (2006), "Improved Sharpe versus Sharpe under the Multiple Index Framework; A Monte-Carlo Simulation Study for the Positive Correlation Case", *Proceedings of 11th International Conference on Industrial Engineering, Theory, Applications and Practice*, Nagoya, Japan, October 24-27, 2006. (Accepted)
- Hsieh J, (1989b), Scintillation camera and multifocal fan-beam collimator used therein *United States Patent 4823017*

- Hurt, N.E., (1991), "Signal enhancement and the method of successive projections", *Acta Applicandae Mathematicae*, **23**, 145-162.
- Jackson, J.E. (1991), *"A User's Guide to Principal Components"*, New York: Springer-Verlag. pp. 365-371.
- Jolliffe, I.T., (1986), *"Principal Component Analysis"*, New-York: Springer-Verlag.
- Jorion, P., (1988), "On Jump Processes in the Foreign Exchange and Stock Markets", *Review of Financial Studies* **1**, 427-445.
- Kalman RE (1960), "A new approach to linear filtering and prediction problems", *Trans ASME J Basic Eng* **82**, 35-45.
- Karpoff, J. (1987), "The Relation Between Price Changes and Trading Volume: A Survey", *Journal of Financial and Quantitative Analysis*, **22**, 109-126.
- Keim, D. and Stambaugh, R., (1986), "Predicting returns in the stock and bond markets", *Journal of Financial Economics*, **17**, 357-390.
- Kendall, M.G., (1953), "The analysis of Economic Time Series, Part I: Prices", *Journal of the Royal Statistical Society, Series A*, **96**, 11-25.
- Kim, M., Nelson, C. and Startz, R., (1991), "Mean reversion in stock prices? A reappraisal of the empirical evidence", *Review of Economic Studies*, **58**, 515-528.
- Kullback, S., and Liebler, R.A (1951), "On information and sufficiency", *Annals of Mathematical Statistics*, **22**, pages 79-86
- Lambert, P., and S. Laurent (2001), *"Modelling Financial Time Series Using GARCH-Type Models and a Skewed Student Density"*, Mimeo, Université de Liège.
- Lamoureux, C.G., and Lastrapes, W.D. (1990), "Persistence in Variance, Structural Change, and the GARCH Model", *Journal of Business and Economic Statistics*, **8**, 225-234
- Lamoureux, C. G., and G. Zhou, (1996), "Temporary components of stock returns: What do the data tell us?", *Review of Financial Studies* **9**:4, Winter, 1033-1059.
- LeRoy, S., (1989), "Efficient capital markets and martingales", *Journal of Economic Literature*, **27**, 1583-1621.

- Lintner, John. (1965), "The Valuation of Risk Assets and the Selection of Risky Investments in Stock Portfolios and Capital Budgets" *Review of Economics and Statistics*, **47**, 13–37.
- Litterman R. and Winkelmann, K. (1998), "Estimating covariance matrices", *Goldman Sachs Risk Management Series*, 1-47
- Ljung, G. M. and G. E. P. Box (1978), "On a measure of lack of fit in time series models", *Biometrika*, **65** p297-303
- Lo, A. and MacKinlay, C., (1988), "Stock market prices do not follow random walks: Evidence from a simple specification test", *Review of Financial Studies*, **1**, 41-66.
- Lucas, R.E, Jr. (1978), "Asset Prices in an Exchange Economy", *Econometrica*, **46**, 1429-1445.
- Lutkepohl, (1991), *"Introduction to Multiple Time Series Analysis"*, Springer Verlag, New York.
- Mallows, C.L., (1973), "Some Comments on Cp", *Technometrics*, **15** (4), 661-675.
- Mandelbrot, B. (1963), "The variation of certain speculative prices", *Journal of Business*, **36**, 394-419.
- Markowitz, H.M., (1952), "Portfolio Selection", *Journal of Finance*, **7**, 77-91
- Markowitz, H.M., (1959), *"Portfolio Selection"*, New York, John Wiley & Sons.
- McQuarrie, A.D.R. and Tsai, C.L., (1998), *"Regression and Time Series Model Selection"*, London: World Scientific Publishing.
- Merton, Robert. C. (1973). "An Intertemporal Capital Asset Pricing Model", *Econometrica* **41**, 867–88.
- Michael, S. Lewis Beck (1993), *"Regression Analysis"*, Sage Publications
- Milhoj A. (1985), "The moment structure of arch processes". *Scandinavian Journal of Statistics*, **12**, 281—292,
- Mills, T. (1999), *"The Econometric Modelling of Financial Time Series"*, Cambridge University Press.
- Mossin, Jan. (1966), "Equilibrium in a Capital Asset Market", *Econometrica* **34**, 768–83.

- McCulloch J H (1985), "Interest-risk sensitive deposit insurance premia: stable ACH estimates", *J. Banking Finance*, **9**, 137–56
- Nelson D B (1990a), "Stationarity and persistence in the GARCH(1, 1) model", *Econometric Rev*, **6**, 318–34
- Nelson, D. B. (1991), "Conditional heteroskedasticity in asset returns: A new approach", *Econometrica* **59**, 347—370.
- Neter, J., Kutner, J. H., Nachtsheim, C. J., and Wassermann, W. (1996), "Applied Linear Statistical Models", Irwin, Chicago.
- Newey, W. K. (1988), "Adaptive estimation of regression models via moment restrictions", *Journal of Econometrics*, **38**, 301–339.
- Osbourne (1959), "Brownian Motion in the Stock Market", *Operations Research*, **7**, 145-73
- Pagan, A., and Ullah A. (1999). "Nonparametric Econometrics", Cambridge University Press, Cambridge.
- Palm, F.C. and P.J.G. Vlaar, (1997), "Simple Diagnostics Procedures for Modelling Financial Time Series", *Allgemeines Statistisches Archiv*, **81**, 85-101.
- Peña, D., Tiao, G.C. and Tsay, R.S., (2001), "A Course in Time Series Analysis", New York: John Wiley & Sons, Inc.
- Philip Hans Franses (1999), "Time Series Models for Business and Economic Forecasting"
- Plaut, G. and Vautard, R., (1994), "Spells of Low-Frequency and Weather Regimes in the Northern Hemisphere", *Journal of the Atmospheric Sciences*, **51**, 210-236.
- Poterba, J. and Summers, L., (1988), "Mean reversion in stock prices: Evidence and implications", *Journal of Financial Economics*, **22**, 27-59.
- Qian, E. and Gorman, S. (2001), "Conditional distribution in portfolio theory", *Financial Analysts Journal*, March/April 2001, 44-51
- Rencher, A. C., (1995), "Methods of Multivariate Analysis", John Wiley & Sons Inc., New York, New York.
- Rencher, A. C. (1998), "Multivariate Statistical Inference and Applications", John Wiley & Sons Inc., New York.

- Richardson, M. and Stock, J., (1989), "Drawing inferences from statistics based on multi-year asset returns", *Journal of Financial Economics*, **25**, 323-348.
- Roberts (1959), "Stock Market "Patterns" and Financial Analysis: Methodological Suggestions", *Journal of Finance*, **14**, 1-10.
- Ross, S.A. (1976), "Arbitrage Theory of Capital Asset Pricing", *Journal of Economic Theory*, **13**, 341-60.
- SAS/STAT User's Guide*, (1990), Version 6, 4th Edition, SAS Institute Inc., Cary, NC (1990).
- Schwarz, G. (1978), "Estimating the Dimension of a Model", *Annals of Statistics*, **6**, 461-464.
- Schwert, G.W. (1989)., "Why does stock market volatility change over time?", *Journal of Finance*, **44**, 1115-53.
- Sharpe, W.F. (1964), "Capital Asset Prices: a Theory of Market Equilibrium under Conditions of Risk", *Journal of Finance* **19**, 425-42.
- Sharpe, W. F. (1970), "*Portfolio Theory and Capital Markets*", McGraw-Hill, New York,
- Sing, T.F. and Ong, S.E. (2000), "Asset allocation in a downside risk framework", *Journal of Real Estate Portfolio Management*, **6:3**, 213-226
- Sparks, R. S.; Coutsourides, D.; and Troskie, L. (1983), "The Multivariate Cp", *Commun. Statistis. Theor. Meth.*, **12** (15), 1775-1793.
- Sparks, R. S.; Zucchini, W.; and Coutsourides, D. (1985), "On Variable Selection in Multivariate Regression", *Commun. Statistis. -Theor. Meth*, **14** (7), 1569-1587.
- Subba Rao, T. (1981), "On the theory of bilinear time series model", *Journal of the Royal Statistics Society*, **43**, No. 2, pp. 244-255.
- Takens, F., (1981), "Detecting strange attractors in turbulence", In D.A. Rand and L.-S. Young (Eds.), 'Dynamical Systems and Turbulence', **volume 898** of *Lecture notes in Mathematics*, 366-381. Springer-Verlag, Berlin.
- Taylor, S. J., (1986). "*Modeling Financial Time Series*", New York: Wiley.
- Tong, T. (1987), "Toward an engineering science of knowledge-based design", *Artificial Intelligence in Engineering*, **2(3)**, 133-166.

- Tong, H. (1990), *"Non-linear time series. A dynamical system approach"*, Clarendon Press, Oxford.
- Trefethen, L.N. and Bau, D. III, (1997), *"Numerical Linear Algebra"*, Philadelphia: SIAM.
- Troskie, C.G., (2000), *STA343S Class notes*, "Principal Components Analysis"
- Troskie, C.G., (2001), *STA403W Class notes*, "Kalman Filter and State Space Models"
- Troskie, C.G., (2002), *STA403W Class notes*, "Multivariate Time Series Modelling"
- Tsay, R. S. (1987), "Conditional heteroscedastic time series models", *J. Am. Statist. Ass.*, **82**, 590–604.
- Vautard, R. and Ghil, M., (1991), "Interdecadal oscillations and the warming trend in global temperature time series", *Nature*, **350**, 324–327.
- Weiss, A. (1986), "Asymptotic Theory for ARCH Models: Estimation and Testing", *Econometric Theory*, **2**, 107–131.
- Working (1934), "A Random-Difference Series for Use in the Analysis of Time Series", *Journal of the American Statistical Association*, **29**, 11–24. (1960), "Note on the correlation of First Differences of Averages in a Random Chain", *Econometrica*, **28**, 916–18.
- Zakoian, J.M. (1990) "Threshold Heteroskedastic Models," manuscript, CREST, INSEE, Paris.
- Zakoian, J.-M (1994), "Threshold Heteroscedastic Models", *Journal of Economic Dynamics and Control*, **18**, 931–995.
- Zhang, P. (1993), "A new multidimensional time series forecasting method based on the EOF iteration scheme", *Advances in Atmospheric Sciences*, **10**, 243–247.